University of Alberta

OIL SAND SCREEN MODELLING USING LINEAR REGRESSION

by

John W. Sheldon

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

Department of Computing Science

Edmonton, Alberta
Fall 2007

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-33349-5
Our file Notre référence
ISBN: 978-0-494-33349-5

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

All truths are easy to understand once they are discovered; the point is to discover them.

– Galileo Galilei (1564 - 1642)

To Mom and Dad,
Who have always supported me, no matter where life has led me.

Abstract

In the oil sands industry, screening is a critical part of the mining process. Syncrude Canada Ltd., an oil sand company, uses screens to separate oversized lumps from the oil-rich sand before it enters the extraction process. However, under unknown screening conditions, some of the sand will pass over the screens, resulting in unexplained variations in screening performance. To investigate these variations, multiple linear regression is used on data from historical databases to identify water and geological variables that affect screening performance. A prediction model, developed using partial least squares regression, is compared to a simple linear model that uses only the oil sand feed rate. Results show an average 25 percent reduction in RMS error over a feed-rate-only model. This is the first known study to identify plant variables, other than the feed rate, that provide insight into oil sand screening behaviour.

Acknowledgements

I would like thank my supervisors, Ronald Kube and Hong Zhang, for their support and encouragement, and for their confidence in me throughout this research. I would like to thank NSERC and Syncrude Canada Ltd. for their financial support, and would like to acknowledge Syncrude Canada Ltd. for providing the necessary water and geological data for this research. Many thanks to Christian West and Brenda Wright at Syncrude Canada Ltd. for their invaluable information regarding the mining process. I am very grateful to the members of my committee, Nilanjin Ray and Ming Zuo, who took the time to read and comment on this work. Special thanks to Dipti Mukherjee for his endless supply of ideas and enthusiasm. Finally, I am very grateful to those individuals in the CIMS lab and the Robotics lab who have helped and supported me throughout this research.

Table of Contents

1 Introduction 1
- 1.1 Oil Sand Mineral Processing . . . 1
- 1.2 Screen Modelling . . . 3
- 1.3 Thesis Objectives and Contributions . . . 4
- 1.4 Methodology and Results . . . 4
- 1.5 Organization . . . 5

2 Background and Related Work 6
- 2.1 Introduction . . . 6
- 2.2 Evaluation of Screen Performance . . . 7
- 2.3 Traditional Screen Variables . . . 9
- 2.4 Traditional Screen Models . . . 12
  - 2.4.1 Empirical Models . . . 12
  - 2.4.2 Probabilistic Models . . . 17
  - 2.4.3 Kinetic Models . . . 19
  - 2.4.4 Evaluation of Traditional Screen Models . . . 24
- 2.5 Summary . . . 26

3 Regression Techniques 27
- 3.1 Introduction . . . 27
- 3.2 Outline of Regression Model-Building Process . . . 27
- 3.3 Statistical Regression Methods . . . 28
  - 3.3.1 Multiple Linear Regression . . . 28
  - 3.3.2 Response Surface Regression . . . 29
  - 3.3.3 Principal Component Regression . . . 30
  - 3.3.4 Partial Least Squares Regression . . . 31
- 3.4 Presentation of Results from Regression Analysis . . . 33
- 3.5 Summary . . . 36

4 Model Development 37
- 4.1 Introduction . . . 37
- 4.2 Plant Screen Variables . . . 38
  - 4.2.1 Water Variables . . . 38
  - 4.2.2 Geological Variables . . . 39
- 4.3 Data Collection . . . 41
- 4.4 Outline of Analysis . . . 50
- 4.5 Results . . . 51
  - 4.5.1 Explanatory Model Development . . . 51
  - 4.5.2 Prediction Model Development . . . 57
- 4.6 Discussion . . . 61
- 4.7 Summary . . . 64

5 Conclusion and Future Work 65
- 5.1 Oil Sand Screen Modelling . . . 65
- 5.2 Future Work . . . 66

Bibliography 67

List of Tables

2.1 Traditional screen design and operating variables. 10
2.2 Common screening factors for screen design and operating variables. The deck position variable indicates the deck level that is being considered [26]. 14
2.3 Standard screen aperture shapes and their respective factor values [26]. 14
2.4 Standard feed conditions and their respective factor values [26]. 14
2.5 List of kinetic variables used in the overview of kinetic screening theory. 20

3.1 NIPALS pseudo-algorithm [29], where N is the number of desired latent vectors. . 32
3.2 Final results using MLR on only the first resulting component from PCR and PLSR respectively. 33
3.3 Example data set for one response variable and five predictor variables. 34
3.4 Resulting performance measures from multiple linear regression performed on the example data set in Table 3.3. 34
3.5 Variables used in the discussion of regression performance measures. 34

4.1 List of water system variables. The first column lists the location in Figure 4.1 where the variable is recorded in the water system. 39
4.2 List of geological variables extracted from the QPD and PI databases. 40
4.3 List of 52 water and geological variables under study. 41
4.4 Algorithm used to calculate the time taken for a truck load of material to move through the surge pile, or the traversal time. The totalTime$_j$ is the traversal time of truck load j. 42
4.5 Periods in Figures 4.4 and 4.5 that are longer than one day. 44
4.6 The number of data points removed at each stage of the data cleaning process for the Summer data set. C1 and C2 refer to crushers NMFB6 and NMFB7 respectively. 49
4.7 The number of data points removed at each stage of the data cleaning process for the Winter data set. C1 and C2 refer to crushers NMFB6 and NMFB7 respectively. 49
4.8 The number of data points for each data set without grouping and the number of data points in each data set when grouped using a 30 minute median filter. The 30 minute median filter required a minimum of 23 data points for each group. Crusher 1 and Crusher 2 refer to crushers NMFB6 and NMFB7 respectively. 49
4.9 Data set naming convention used in the data analysis portion of this work. 50
4.10 List of the significant variables for summer and winter screening conditions. . . . 51
4.11 Models used to determine how significant the identified summer screen variables are for summer screening conditions. The first model, SLR, is a simple linear regression model using only the feed tonnage as input. The second model, MLR$_{sig}$, uses only the summer significant variables in Table 4.10. The third model, MLR$_{all}$, uses all water and geological variables. The last model, MLR$_{comb}$, a model that uses a common set of variable coefficients for both the Train 1 and Train 2 data sets. It is used to determine whether a combined model for both Train 1 and Train 2 is possible. 52
4.12 Models used to determine how significant the identified winter screen variables are for winter screening conditions. The first model, SLR, is a simple linear regression model using only the feed tonnage as input. The second model, MLR$_{sig}$, uses only the winter significant variables in Table 4.10. The third model, MLR$_{all}$, uses all water and geological variables. The last model, MLR$_{comb}$, a model that uses a common set of variable coefficients for both the Train 1 and Train 2 data sets. It is used to determine whether a combined model for both Train 1 and Train 2 is possible. 53

4.13 The results of two regression models that use two different sets of variables on the Summer Train 2 data set. The average overflow tonnage during the period is 1076 TPH. The SLR model uses only the feed tonnage as input variable. The MLR model uses the identified significant geological variables, including the feed tonnage variable, listed in Table 4.10. The pumpbox valve percent open variable is not included in order to underline the importance of the geological variables. 54
4.14 The results of two regression models that use two different sets of variables on the Winter Train 2 data set. The average overflow tonnage during the period is 855 TPH. The SLR model uses only the feed tonnage as input variable. The MLR model uses the identified significant water variables listed in Table 4.10. 55
4.15 Coefficient values and statistical evaluation measures for the summer screen submodel. 56
4.16 Coefficient values and statistical evaluation measures for the winter Train 1 screen submodel. 56
4.17 Coefficient values and statistical evaluation measures for the winter Train 2 screen submodel. 56
4.18 A list of regression methods used to develop prediction models. Each regression method is run twice, a first time on the summer or winter significant variables, and then a second time with all 52 water and geological variables. The response surface regression method is not run on all 52 variables due to computational restrictions resulting from the large number of predictor variables. 58
4.19 Method used to calculate the relative performance between the best performing model and the remaining models. M is the total number of randomly selected groups of data in the cross-validation and b is the best performing model. 58
4.20 Results for each model for the Summer Train 1 data set. The relative performance between the best performing model and the remaining models is provided in the last two columns. The best performing model for the Summer Train 1 data set is the PLSRa model. The overflow tonnage for the Summer Train 1 data set is 1142±375 TPH. 59
4.21 Results for each model for the Summer Train 2 data set. The relative performance between the best performing model and the remaining models is provided in the last two columns. The best performing model for the Summer Train 2 data set is the PLSRa model. The overflow tonnage for the Summer Train 2 data set is 1076±340 TPH. 59
4.22 Results for each model for the Winter Train 1 data set. The relative performance between the best performing model and the remaining models is provided in the last two columns. The best performing model for the Winter Train 1 data set is the PLSRa model. The overflow tonnage for the Winter Train 1 data set is 1086±469 TPH. 60
4.23 Results for each model for the Winter Train 2 data set. The relative performance between the best performing model and the remaining models is provided in the last two columns. The best performing model for the Winter Train 2 data set is the PLSRa model. The overflow tonnage for the Winter Train 1 data set is 855±279 TPH. 60
4.24 Variable coefficients for the Summer screen model, the Winter Train 1 screen model, and the Winter Train 2 screen model. 61
4.25 Facies 8 and 21 descriptions. 62
4.26 List of the four regression methods used to develop the oil sand prediction screen models. 63

List of Figures

1.1 Example of a magnified oil sand particle with surrounding layers of water and bitumen. 2
1.2 Syndrude Canada Ltd. North Mine oil sand comminution and screening system. . 2
1.3 A profile view of the static screens in use at Syncrude Canada Ltd. North Mine. In this work, reject oil sand material is referred to as overflow material. 3

2.1 Example of a typical efficiency curve and an ideal efficiency curve. 8
2.2 Example of the observed relationship between the screening efficiency and the feed tonnage. 11
2.3 Cross-section of a screen aperture showing the effective aperture used in the Karra model [24]. 15
2.4 A typical overflow partition curve for a vibrating screen. 17
2.5 Graph of Gaudin's overflow partition curve for N = 5 and N = 15. The aperture size is 55 mm. Note the similar shape to that of Region B in Figure 2.4. 18
2.6 The passage rate as a function of the position along the screen. Also marked are the crowded and separated regions on the screen. 20
2.7 The stratification of particles within a bed of material. 23
2.8 A gray coded scatter plot of overflow tonnage versus feed tonnage during summer months (May to August) for 2005 and 2006 for the screening system in Figure 1.2. The varying levels of gray represent the local data point density ranging from zero (white) to 30 (black). 25
2.9 A gray coded scatter plot of overflow tonnage versus feed tonnage during winter months (December to March) for 2005 and 2006 for the screening system in Figure 1.2. The varying levels of gray represent the local data point density ranging from zero (white) to 30 (black). 26

3.1 Regression model-building process [37]. 28
3.2 Example of principal components in a two-dimensional data set. 30
3.3 Example values for a response vector, Y, and two predictor variables, X_1 and X_2. . 32
3.4 Values of the response and predictor variables after centring and scaling. 32
3.5 Plot of centred and scaled example data with overlaid PCR components. 33
3.6 Plot of centred and scaled example data with overlaid PLSR latent vectors. 33

4.1 Diagram of water flow in the screening system. 39
4.2 The extraction of a core sample and identification of different facies. The percentage value indicates the percentage of the total volume that facies occupies. 40
4.3 Surge pile with two crusher conveyor belts depositing material onto the surge pile and two feed conveyor belts removing material from underneath. Two laser range finders are located on the exit end of each crusher conveyor belt to measure the height of the surge pile. 43
4.4 Plot of the time difference in minutes between the PI data point and the nearest QPD ore data point for the Summer data set. 44
4.5 Plot of the time difference in minutes between the PI data point and the nearest QPD ore data point for the Winter data set. 44
4.6 Histogram of the period lengths between a data point recorded at one crusher and a data point recorded at the second crusher for the Summer data set. Period lengths greater than 15 minutes are placed in the 15 minute bin. 45
4.7 Histogram of the period lengths between a data point recorded at one crusher and a data point recorded at the second crusher for the Winter data set. Period lengths greater than 15 minutes are placed in the 15 minute bin. 45
4.8 Histogram of the status flag values for the Summer data set. 46
4.9 Histogram of the status flag values for the Winter data set. 46

4.10 Histogram of the overflow tonnage values for the Summer data set. 47
4.11 Histogram of the overflow tonnage values for the Winter data set. 47
4.12 Histogram of the feed tonnage values for the Summer data set. 47
4.13 Histogram of the feed tonnage values for the Winter data set. 47
4.14 Histogram of the temperature values for the cleaned Summer data set. 48
4.15 Histogram of the temperature values for the cleaned Winter data set. 48
4.16 Model overflow tonnage versus the actual overflow tonnage for the Winter Train 1 data set. 53
4.17 Model overflow tonnage versus the actual overflow tonnage for the Winter Train 2 data set. 53
4.18 Model overflow tonnage versus the actual overflow tonnage for the SLR model for the Summer Train 2 data set. 54
4.19 Model overflow tonnage versus the actual overflow tonnage for the MLR model for the Summer Train 2 data set. 54
4.20 Model overflow tonnage versus the actual overflow tonnage for the SLR model for the Winter Train 2 data set. 55
4.21 Model overflow tonnage versus the actual overflow tonnage for the MLR model for the Winter Train 2 data set. 55

Chapter 1

Introduction

Separating material has preoccupied humans for thousands of years. The action of separation has extended into such fields as agriculture [56], mineral processing [36, 26], and more recently, sorting of recycled household products [46, 40]. In oil sand mineral processing, material separation is accomplished using large static and vibrating screens.

The current understanding of oil sand screening phenomena is lacking despite the growing interest in optimizing the screening process. This forces screen installation engineers and operation personnel to consider theory from the hard rock mineral processing industry. Unfortunately, hard rock material is geologically different from oil sand material and it is unknown if hard rock models can be applied to oil sand screens. For this reason, there exists a strong desire for a greater understanding of oil sand screening phenomena and those factors that affect screen performance.

The goal of this research is to identify water and geological variables that exhibit a strong relationship with screen performance. From the identification of important variables, explanatory and predictive screen models can be developed. Ultimately, a greater understanding of the screening system and those factors that affect screen performance can be acquired.

1.1 Oil Sand Mineral Processing

Oil sand mineral processing in Northern Alberta, Canada is currently growing at a phenomenal rate. At the heart of the industry is a material referred to as *oil sand*. Oil sand is a mixture of bitumen, sand, clay, and water, and its physical properties are found to vary from season to season. In warm summer conditions, oil sand is thick and sticky resembling a mixture of molasses and rocks. In colder winter conditions, oil sand freezes creating material that resembles hard rock. At the heart of oil sand are oil sand particles, shown in Figure 1.1, that contain a layer of water sandwiched between a particle of sand and a layer of bitumen. The layer of water enables the bitumen to be easily removed from the oil sand using water-based extraction methods [27]. Once the bitumen is removed from the sand particle it is processed to form products like oil and gas.

Syncrude Canada Ltd., currently one of the largest oil sand companies in the world, operates

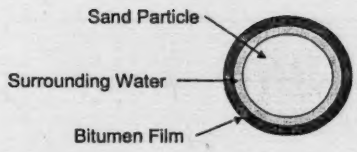

Figure 1.1: Example of a magnified oil sand particle with surrounding layers of water and bitumen.

an oil sand extraction facility in Northern Alberta known as North Mine [27]. A schematic of the comminution and separation system is provided in Figure 1.2.

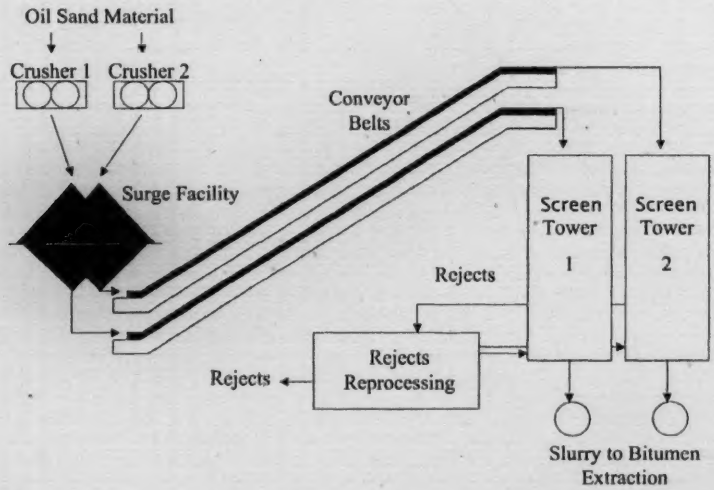

Figure 1.2: Syndrude Canada Ltd. North Mine oil sand comminution and screening system.

Syncrude Canada Ltd. processes, on average, between 10 and 12 thousand tonnes per hour of oil sand material in their North Mine mineral processing facility. Oil sand is first placed into dump trucks using either hydraulic or cable shovels and transported to large toothed roll crushers. Once the material has been crushed, it passes through a surge facility and is dumped onto conveyor belts that transport the oil sand to one of two screen towers. Each screen tower contains a series of large static screens, shown in Figure 1.3, that remove material too large to be handled in the subsequent extraction process. Material that passes over the static screens is transported to a reprocessing system where it is again crushed and screened using an impact hammer and a vibrating screen. Material that fails to pass through the vibrating screen is placed in a dump truck and either returned to the crushers or marked as waste material. Material that does pass through the static screens or the rejects reprocessing vibrating screen is mixed with hot water to form a slurry. The slurry is then pumped to

an extraction facility where bitumen is removed from the sand particles.

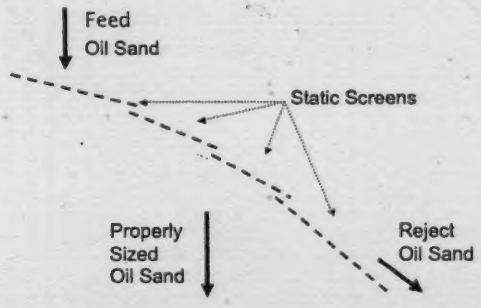

Figure 1.3: A profile view of the static screens in use at Syncrude Canada Ltd. North Mine. In this work, reject oil sand material is referred to as overflow material.

The static screens play a critical role in the oil sand mining process. The static screens ensure a fixed maximum particle top size in the material allowing the process downstream to be operated effectively. However, during typical screen operation, poor operating conditions and complex particle screen dynamics cause oil sand particles to be rejected when they should have passed through. This creates variations in the screen overflow tonnage that are not explained by the feed tonnage alone. To address this problem, a static screen model is needed to understand and predict variations in screen overflow tonnage.

1.2 Screen Modelling

A system model relates a series of input variables to one or more output variables. In the mining industry, comminution and screening systems are commonly modelled using a black box model. Black box models create relationships between input variables and output variables using large data sets. Examples of screen input variables are the feed particle size distribution and the feed tonnage. Common output variables are the overflow tonnage and the screen efficiency.

Screen modelling research has never been carried out in detail for large static screens similar to those in operation at Syncrude. While a screen model may not be available for the specific type of screen considered in this work, existing screen models for other types of screens may adequately capture the behaviour of the system. When considering all available industrial screens, industrial aggragate vibrating screens are similar to oil sand static screens in terms of both the size range of the particles screened and screen's physical characteristics. Further discussion of the similarities between static and vibrating screens are discussed in Chapter 2 along with three major areas of vibrating screen models.

Screen modelling has been well researched in the hard rock industry. However, accurate prediction of screen performance in an operational plant is often not realized from existing screen

models. Inconsistent screen model performance is attributed to the large number of variables, such as the particle size distribution, that can influence particle motion. Since existing models are difficult to implement and may be inaccurate, plant designers commonly rely on their own background knowledge or trial and error experiments to select screen specifications. This approach can increase dramatically the time and cost investment of a screen installation. Furthermore, due to the absence of accurate screen models, screen operation becomes reliant on operator experience resulting in suboptimal screen performance. This is also true for the oil sand industry and motivates the development of an oil sand screen model. This research is the first step towards an oil sand screen model that will enable designers to select screen specifications with confidence and allow screen operators to make informed decisions during screen operation, ultimately resulting in an increase in screen performance.

1.3 Thesis Objectives and Contributions

The work presented in this thesis focuses on identifying significant water and geological variables and developing a predictive oil sand screen model. Currently, overflow tonnage variation is modelled using only the feed tonnage, with often poor results. We test the hypothesis that through the identification of significant water and geological variables, a screen model can be developed that allows a stronger understanding of overflow tonnage variation. The goal is to show that water and geological variables will provide a greater explanation and prediction of screen performance variation than feed tonnage alone.

A number of research contributions are made in this thesis to the field of oil sand screen modelling. We provide a general overview of traditional approaches to screen modelling and justify the decision to take an empirical approach to designing an oil sand screen model using historical plant data. We identify significant water and geological variables that show strong correlations with the overflow tonnage. Furthermore, the identified significant variables validate claims made by operators concerning their significance. We develop an explanatory model using the identified significant variables to provide insight into screen performance variance for both summer and winter screening conditions. Finally, we develop predictive models that are shown, using cross-validation, to perform significantly better than current oil sand screen models that use only the feed tonnage. The result of this work is a greater understanding of the screening process at Syncrude's oil sand mineral processing facility.

1.4 Methodology and Results

To test the hypothesis presented in this work, oil sand geological variables, such as the ore grade, and water variables, such as the water flow, are collected for eight months of summer operation and eight months of winter operation over two years. The data is cleaned and grouped into 30 minute

periods to account for uncertainty in the data. From the original set of variables, significant variables are identified for both summer and winter screening conditions using forward stepwise multiple linear regression to gain insight into the oil sand screening process. The significant variables are then used to build a preliminary explanatory screen model using multiple linear regression. The coefficients of the explanatory model are analyzed to understand the relationship of each variable with the overflow tonnage. Finally, multiple linear regression, response surface regression, principle component regression, and partial least squares regression are used to develop a prediction model. Each regression method is run a first time on either the summer or winter significant variables and then a second time on all variables. An exception is the response surface regression method, which is not run on all variables due to computational restrictions resulting from the large number of predictor variables. Each resulting model was evaluated, using five-fold cross-validation and standard error analysis methods, to select an oil sand screen model that provides the greatest predictive accuracy.

In this work, important variables are identified for both summer and winter screening conditions that are shown to be statistically significant. Multiple linear regression is used to develop explanatory models for both summer and winter months. It is determined that a noticeable difference in Train 1 and Train 2 screen performance exists during the winter screening period. A predictive model developed using partial least squares regression provided the best performance, with an average reduction in RMS error of 25 percent over a simple feed-only model. The analyses were carried out using data representing both summer and winter conditions over a two year period ensuring that the results are statistically significant.

1.5 Organization

The organization of the remainder of the thesis is as follows. Chapter 2 provides an overview of existing screen models and identifies limitations that motivate the development of a new static screen model using statistical regression methods. Chapter 3 outlines the methodology of building a regression model, presents four common statistical regression methods, and introduces several typical regression performance measures. Chapter 4 discusses data collection and cleaning of historical data and the identification of significant water and geological variables for summer and winter screening conditions. In addition, four regression methods are used to develop oil sand prediction models that are compared to a simple screen model that uses only the feed tonnage as input. Finally, results are presented followed by a discussion of the various insights gained into oil sand screen operation. Chapter 5 provides a summary of the work and a number of possible future directions for oil sand static screen modelling research.

Chapter 2

Background and Related Work

In the previous chapter we introduced background information on oil sand mineral processing at Syncrude Canada Ltd. and the research problem addressed in this thesis. In this chapter we present background and related work in the area of oil sand screen modelling. Finally, we present motivation for the development of a new static screen model using an empirical approach.

2.1 Introduction

Syncrude's oil sand static screens are among the largest in the world at over two metres in width and eight metres in length. However, due to their relatively recent introduction into oil sand screening, public research is unavailable. Before undertaking the development of an oil sand specific empirical static screen model, background literature on common aggragate industrial screens are presented followed by an overview of industrial hard rock screen models.

General particle separation systems can be classified based on the size of the material to be separated and the medium in which the particles are transported. For material transported in water, hydrocyclones are used to separate fine material [36] and sieve bends allow dewatering and separation of larger material [28]. For dry screening of particles below 50 microns air cyclones can be used [34]. For dry material above 50 microns, vibrating and static screens are typically used [62]. For particles above several hundred millimetres, static and vibrating grizzly screens are commonly used [36]. For general hard rock screening of material between several to several hundred millimetres, similar to material in the oil sand industry, vibrating screens such as banana screens, centrifugal screens, and shaking screens are commonly used.

When considering each of the above screen types, vibrating screens are found to offer the greatest similarity to oil sand static screens. The particle size range and volume of material processed is similar in both cases. Furthermore, vibrating screen surfaces are typically either wire mesh, perforated steel, or polyurethane, which are similar to Syncrude's oil sand static screens. A final advantage of using vibrating screen models is the existence of publicly available research that can be quickly evaluated on data from Syncrude's static screens. In the following sections, screen performance

evaluation methods, typical screening variables, and common hard rock vibrating screen models are presented. In the last section we address the applicability of hard rock screen models to oil sand screens.

2.2 Evaluation of Screen Performance

The mineral processing industry has developed two commonly used means of measuring screen performance: efficiency measures and capacity measures. Efficiency measures evaluate how well the screen is separates the material into two size categories. Two efficiency measures are efficiency indices and efficiency curves. Capacity measures evaluate the quantity of material a screen can separate effectively.

Efficiency Indices

Representing the screen performance by a single value is the simplest way to convey performance information. Efficiency indices range from zero to one, where a higher value typically indicates greater screening efficiency. A commonly used efficiency index, shown in Equation 2.1, calculates the ratio of material that actually passed through the screen to the material that should have passed through the screen [22, 34].

$$\text{Efficiency} = \frac{\text{F - O}}{\text{F - O} \cdot o_o} \tag{2.1}$$

where F is the feed tonnage, O is the overflow tonnage, or tonnage of material that passes over the screen, and o_o is the percentage of oversize material in the overflow tonnage, where oversize material are particles larger than the screen aperture. This efficiency index assumes that oversize particles do not pass through the screen. However, if the screen contains worn or oversized apertures, Equation 2.1 will not capture the resulting decrease in screen efficiency due to misplaced oversize particles.

Alternative indices include an index proposed by Hess [22] which accounts for particles whose size are approximately equivalent to the screen aperture. These particles were shown to reduce screen efficiency by requiring a longer time to pass through the screen. In other work, Apling [1] proposed Equation 2.2, which accounts for both the percentage of oversize material in the underflow, u_o, and the percentage of undersize material in the overflow, o_u. As the percentage of misplaced material decreases resulting from better material separation, the efficiency of the system increases.

$$\text{Efficiency} = 1 - u_o - o_u \tag{2.2}$$

Alternatively, Partridge [42] proposed a screen efficiency index which is equal to the percentage of undersize in the underflow minus the percentage of oversize in the underflow. This index penalizes for both misplaced undersize and misplaced oversize particles.

Efficiency indices assume that the desired cut size is the size of the aperture, which results in a negligible probability of misplaced oversize particles. This is not the case in situations where the cut size is less than the aperture size or when the particles are irregularly shaped. Ultimately, efficiency indices are typically developed towards a specific application. Consequently, transferring an efficiency index from one type of screen to another may result in unsatisfactory evaluation of screen performance [22].

Efficiency Curves

A common concern in screen performance evaluation is that a single value does not provide sufficient information of the actual material separation [22]. Efficiency indices do not capture the particle size distribution of the overflow or underflow material that gives information of the quality of separation.

An alternative to using an index is to use an efficiency curve similar to the one in Figure 2.1. Efficiency curves are constructed from the percentage of particles that pass over the screen for each particle size bin. Efficiency curves are also referred to as oversize partition curves or overflow partition curves.

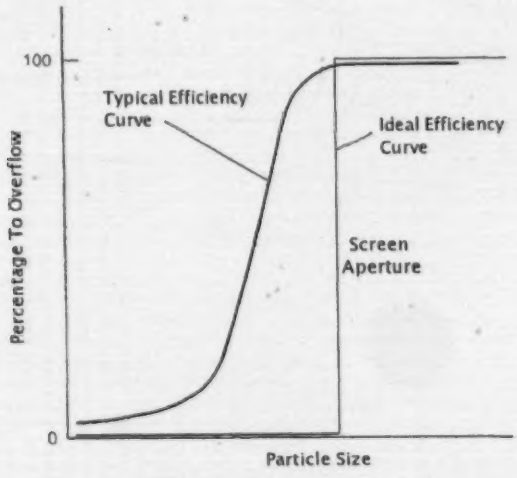

Figure 2.1: Example of a typical efficiency curve and an ideal efficiency curve.

Three well known distribution functions that successfully describe the shape of the efficiency curve are the Rosin-Rammler distribution [45], the Weibull distribution [11], and the Gaudin-Schuhmann distribution [47]. The Rosin-Rammler distribution, shown in Equation 2.3, was first used to describe the overflow particle size distribution, $P(x, \alpha)$, of powdered coal after crushing.

$$P(x, \alpha) = 1 - exp\left[-\left(\frac{x}{x_{63.2}}\right)^{\alpha}\right] \tag{2.3}$$

where x is the particle size, typically measured as the diameter of a sphere with equivalent volume, and $x_{63.2}$ is the bin size where 63.2 percent of the particles have a diameter less than the aperture. The Weibull distribution [11] was first proposed to model the fracture of materials under repetitive stress and was later used to describe screen efficiency curves. The functional form of the Weibull distribution is

$$P(x, \alpha, \beta) = 1 - exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] \tag{2.4}$$

where x is the particle size, β is a characteristic size, typically assumed to be the aperture size, such that 63.2 percent of the particles will pass, and α is a fitting constant. The third commonly used size distribution function is the Gaudin-Schuhmann equation [47] shown in Equation 2.5.

$$P(x, k, m) = \left(\frac{x}{k}\right)^{m} \tag{2.5}$$

where x is the particle size, k is a size modulus, and m is the distribution modulus. The parameter values in all three size distribution functions are determined from fitting the distribution function to collected overflow particle size distribution data. The advantage of describing the efficiency curve as a function is that performance information can be obtained for any bin size without the need for imprecise linear interpolation between known values.

Screen Capacity Measure

The second screen performance measure is based on the tonnage of material a screen can separate effectively and is referred to as the capacity measure. In the mineral processing industry, often the goal is to process as much material as possible and as quickly as possible. The capacity measure is then simply the tonnage of material processed effectively by a screening system, where a higher tonnage is better.

However, the capacity measure is inversely related to the efficiency measure making it difficult to maximize both the efficiency measure and the capacity measure. Any increase in feed tonnage will decrease the screen efficiency until the majority of material is passing over the screen to overflow. Therefore, screening is a delicate balance between maximizing both screening efficiency and screen capacity. In the mineral processing industry, the capacity measure is typically given the higher priority where higher capacities result in a greater amount of processed material.

2.3 Traditional Screen Variables

A large number of input screening variables have been used in hard rock screening models [51, 10, 2]. They are traditionally divided into two areas: design variables and operating variables [7]. Table 2.1 lists the screen variables associated with each area.

Design Variables	Operating Variables
Screen Length and Width	Feed Particle Size Distribution
Aperture Size and Shape	Bulk Density
Percentage of Open Area	Feed Tonnage
Screen Angle	Screen Motion
Screen Media	Material Moisture Content
	Wet Screening
	Particle Shape

Table 2.1: Traditional screen design and operating variables.

Design variables are addressed during the design phase of screen development. Once the screen is installed, and during screen operation, design variables remain relatively constant. Operating variables, however, cause screen performance to vary during operation. Traditionally, operating variables are measured and data is collected using sampling techniques [26]. Due to the complex nature of the mining process, in both summer and winter conditions, the number of samples needed is beyond the scope of this work. In essence, data collected over a long period of time on a continual basis, or time series data, is needed to capture the dynamic nature of the screening system. We turn instead to historical process data collected by Syncrude Canada Ltd. and placed into a Process Information (PI) database. Syncrude's PI database is a centralized historical process data collection and storage system that allows staff and researchers to easily access data collected throughout the plant [41]. The historical time series data contained within PI is collected on a continual basis throughout the year and allows the questions posed in this research to be addressed. The following sections review traditional operating variables and identify whether they are measured by Syncrude and placed into the PI database. Note that the screen motion variable listed in Table 2.1 is not discussed further as it does not apply to static screens.

Feed Particle Size Distribution

The feed particle size distribution is an important screen variable for modelling screen performance. Researchers often describe the particle size distribution using three characteristic size ranges. The first size range consists of particles having a size greater than the aperture size and are referred to as oversize particles. A higher percentage of oversize particles enables the screen to handle higher capacities as well as unplugging the screen by impacting lodged particles [51]. The second range consists of particles having a size less than half the aperture size and are referred to as half-size particles. A higher percentage of half-size particles is advantageous since half-size particles easily flow through the screen. The last range consists of particles having a size 75 to 125 percent of the aperture size and are referred to as near-size particles. Near-size particles are known to increase plugging and require much longer screening times to be sized correctly [42]. Currently, at Syncrude's North Mine, time series data is not available for the feed particle size distribution.

Bulk Density

The bulk density is the density of material when it is not compacted. It has been shown that material with a higher bulk density is heavier and tends to fall through the screen more easily, while material with a lower bulk density is lighter and tends to bounce further along the screen [26]. Currently, at Syncrude's North Mine, time series data is not available for the material bulk density.

Feed Tonnage

The feed tonnage is critical to the performance of the screen [51]. Too low a feed tonnage and the particles bounce quickly over the screen without settling. Too high a feed tonnage and particles will have difficulty traversing through the bed of material to get to the screen's surface. Experiments have found that screen efficiency, as a function of feed tonnage, has a parabolic shape similar to the one in Figure 2.2 [60]. Currently, at Syncrude's North Mine, time series data is available for the feed tonnage.

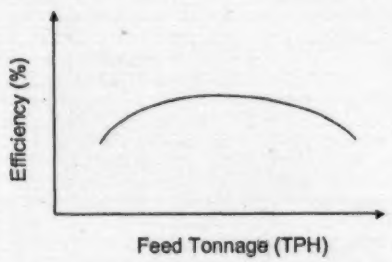

Figure 2.2: Example of the observed relationship between the screening efficiency and the feed tonnage.

Moisture Content

Moisture content refers to the amount of water contained within or surrounding each particle. Higher moisture levels are found to increase particle adhesiveness, resulting in increased plugging and blinding in the screening media [34, 43]. An increase in adhesiveness also results in smaller particles sticking to oversized particles and being carried to overflow. To alleviate the affects of high moisture content, pre-drying the material or the use of spray bars to push material through the screens, also known as wet screening, are typically used [42]. Currently, at Syncrude's North Mine, time series data is not available for the material moisture content.

Wet Screening

Wet screening, as previously mentioned, is used when moisture is present in the material. Wet screening uses high pressure spray bars to push material through the screen. This approach works well for small particles that tend to attach to larger particles and pass to overflow [15]. Syncrude

uses high pressure spray bars to aid the screening process, however, existing models that contain wet screening [13, 31, 20, 39, 15] only consider its influence for aperture sizes less than 50 millimetres. The static screens considered in this work contain aperture diameters of approximately 57 millimetres. Therefore, wet screening is not considered significant, or perhaps was never considered, by existing screen models for the aperture sizes used in oil sand screening. Currently, at Syncrude's North Mine, time series data is not available for wet screening.

Particle Shape

Research has shown that the shape of the particles can affect screen performance where elongated or slab-like particles are harder to screen than spherical particles [34, 26]. Slab-like particles have a higher chance of skipping over screen apertures since particles tend to lie on their axis of greatest stability. Currently, at Syncrude's North Mine, time series data is not available for the material particle shape.

To summarize, in this section traditional vibrating screen variables were reviewed and it was determined that historical time series data is only available for the feed tonnage variable. The remaining variables are not currently measured during plant operation and would require a large number of samples to be taken to obtain the necessary data to investigate their affect on oil sand screen performance.

2.4 Traditional Screen Models

In Section 2.1, different screen types were presented and it was determined that large aggregate vibrating screens showed the greatest similarity to oil sand static screens. Vibrating screens have been studied since as early as the 1930's in order to increase screening efficiency [19]. Since then, proposed screen models have typically fallen into three major categories: empirical models, probabilistic models, and kinetic models. For each category, an overview of the theory is presented below. In the last section, the applicability of each category towards the modelling of oil sand screens is presented.

2.4.1 Empirical Models

Empirical models use regression methods to extract relationships between input variables and output variables from large data sets. They are typically system specific and require no inherent knowledge of the screening process. Empirical models are often developed when existing models perform poorly or when none is available for the system of interest. There are three main groups of empirical models. The first group contains capacity models that use empirically derived screening factors to relate screening variables to the screen capacity. The second group contains efficiency models that fit efficiency curves to collected screen data. The last group of empirical models use advanced statistical regression methods on large data sets and places no assumptions on the underlying form.

Advanced statistical regression empirical methods often create more accurate models of the systems of interest, however, the resulting models typically provide little insight into the underlying system.

Capacity Models

Capacity models are developed by screen manufacturers using designed experiments and are typically screen specific [13, 15, 31]. They are used to evaluate the screen size needed based on the screening conditions and the desired feed tonnage. Screen operators can also use the capacity model to control the amount of feed entering their screening system if the screening conditions change. Therefore, capacity models consider both the screen design and operating variables.

A capacity model calculates the rated capacity of a screen. The rated capacity is the feed tonnage that the screen can handle under a set of screening conditions. The screening conditions are calculated from the screening variables, presented in Section 2.3, through a series of screening factors K_i. During typical operating conditions the screening factors are set to one and the rated capacity is called the based capacity, I_u. Typical operating conditions are those conditions that the manufacturer assumes the screen will be operating in the majority of the time. The rated capacity is calculated from the base capacity and the screening factors using Equation 2.6.

$$\text{Rated Capacity} = I_u \prod_i K_i \tag{2.6}$$

where i is the number of screening variables. As mentioned earlier, during typical screening conditions the screening factors are set to one. Unfortunately, screening conditions are often not typical. When screening conditions do change, the screening factors attempt to capture the effect a change in a screening variable has on the actual screen capacity. For example, if a screening variable increases the actual screen capacity then the value of the factor becomes greater than one. Similarly, if a screening variable decreases the actual screen capacity, then the factor becomes less than one. The screening factors attempt to affect the rated capacity in the same manner the screening variables affect the actual screen capacity. Table 2.2 lists typical screening factors and base capacity calculations [26].

Capacity models are generally used in the mineral processing industry because they are easy to understand and simple to use. However, their simplicity often results in poor performance and, consequently, the models are often modified based on user background knowledge and experience of the screening system.

Efficiency Models

King [26] proposed a screen model that relates the rated capacity of the screen to the screen efficiency. King defines a new value called the rating ratio (RR), which is simply the rated capacity, presented in the previous section, divided by the feed tonnage. Using a large data set, a model relating the rating ratio to the screen efficiency is developed (see Equation 2.7).

Factor Name	Equation
Aperture Size	h
Base Capacity	$I_u = 20.0h^{0.33} - 1.28$ for $h < 25$ mm
	$I_u = 0.783h + 37$ for $h \geq 25$ mm
Open Area	K_1 = (Screen Percentage Open Area)/50
Half-Size	$K_2 = 2 \cdot (\%< 0.5h) + 0.2$
Oversize	$K_3 = 0.914 \exp(\exp(4.22(\% > h) - 3.50))$
Bulk Density	K_4 = Bulk Density/1600
Deck Location	$K_5 = 1.1 - 0.1 \cdot$ Deck Position
Screen Angle	$K_6 = 1.0 - 0.01$(Screen Angle-15 degrees)
Wet Screening	$K_7 = 1.0$ for $h > 25$ mm
Aperture Shape	See Table 2.3
Particle Shape	For spherical particles, K_9=1.0
	For feed with about 15% elongated particles, K_9=0.9
Surface Moisture	See Table 2.4

Table 2.2: Common screening factors for screen design and operating variables. The deck position variable indicates the deck level that is being considered [26].

Shape of Screen Opening	K_8
Round	0.8
Square	1.0
2 to 1 rectangular slot	1.15
3 to 1 rectangular slot	1.20
4 to 1 rectangular slot	1.25

Table 2.3: Standard screen aperture shapes and their respective factor values [26].

Condition of Feed	K_{10}
Wet, muddy or sticky material	0.75
Wet surface quarried and material from surface stock piles with up to 15 percent moisture by volume	0.85
Dry crushed material	1.00
Naturally or artificially dried material	1.25

Table 2.4: Standard feed conditions and their respective factor values [26].

$$\begin{aligned} \text{Efficiency} &= 0.95 - 0.25(RR - 0.8) - 0.05(RR - 0.8)^2 \text{ for RR} \geq 0.8 \\ &= 0.95 - 1.67(RR - 0.8)^2 \text{ for RR} < 0.8 \end{aligned} \tag{2.7}$$

The model presented by King is simple to implement assuming that a good knowledge of the screening variables and their relationship with the rated capacity is available.

Karra [24] also developed an efficiency screen model that relates the rated capacity and the feed particle size distribution to the screen efficiency, $E(x)$. Equation 2.8 was determined by fitting a distribution function to a large data set using statistical regression.

$$E(x) = 1 - exp\left[-0.693\left(\frac{x}{x_{50}}\right)^c\right] \tag{2.8}$$

where x is the feed particle size, c is a constant set at 5.9, and x_{50} is the feed particle size that has a 50 percent chance of passing through the screen. The value of x_{50} is calculated using

$$x_{50} = h_t \left(\frac{K}{G}\right)^{-0.148} \tag{2.9}$$

where h_t is the effective aperture, G is the near-size factor, and K is the rating ratio of the screen given the current screening conditions. Referring to Figure 2.3, the effective aperture is calculated as $h_t = (h + d)cos\theta - d$, where h is the current aperture, d is the width of the wire, and θ is the screen angle.

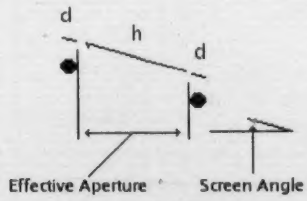

Figure 2.3: Cross-section of a screen aperture showing the effective aperture used in the Karra model [24].

The near-size factor G is calculated using Equation 2.10 where X_n is the percentage of near-size material and a' and c' are regression constants determined to be 0.844 and 3.453 respectively for the test screen.

$$G = a' \left(1 - \frac{X_n}{100}\right)^{c'} \tag{2.10}$$

The model proposed by Karra works well for the data range contained in the data set. Unfortunately, using the model outside of this data range may result in large inaccuracies [36]. Hess [22] notes that the overflow partition curve has the unfortunate property of always passing through the origin and so can cause difficulty when fitting at low particle sizes. The value of the constant in Equation 2.8 was originally kept static, however Hess found better fitting when varying both x_{50} and c. Finally, the Karra model is easy to implement and performs well if the screening conditions are similar to those found in Karra's experimental data set.

Alternative efficiency models proposed by Hatch and Mular [21] and Batterham et al. [4] require only the feed particle size distribution and the screen aperture size. The models use regression to fit distribution functions, used to describe the efficiency curve, to collected data. The models are accurate until screening conditions change, at which time new data must be collected and the models refit.

Advanced Statistical Regression Models

Several researchers in the field of industrial screening developed models purely from data without imposing constraints on the model's functional form. Rongguang et al. [12] used a commercial software package to find correlations between input and output variables using multiple regression analysis. In his work, data was collected from 172 controlled runs that allowed adequate exploration of the range of variables and their influence on the response variables. Their approach was to find all combinations of powers, fractional powers, and cross-products of primary variables, keeping only those that contributed significantly to the model. The authors provided no details in the paper as to how the significant independent variables were identified. Good correlation was noted between their regression results and general understanding of screening behaviour. The authors compared their results to a Monte Carlo probability screen simulation conducted earlier by Beeckmans and Jutan [5] and the level of agreement was found to be satisfactory.

Van der Walt et al. [57, 58] successfully modelled a hydrocyclone classifier using both sigmoidal back-propagation neural networks (SBNN) and multiple adaptive regression splines (MARS). The data was first scaled linearly and logarithmically to ensure proper behaviour when used in the SBNN's and scaled linearly for MARS. Several training runs were conducted for SBNN to determine the best number of hidden nodes. Model selection was based on the normalized absolute error and it was found that the SBNN had the best performance since the system was nonlinear.

Karr and Weck [23] used fuzzy linguistic models with genetic algorithms to model hydrocyclone circuits. The hydrocyclone was characterized by the split size defined as the size where a particle has an equal chance of passing through the hydrocyclone or passing over. A series of eight hydrocyclone parameters, including the diameter and the height of the hydrocyclone, were related to the split size using fuzzy linguistic methods. Genetic algorithms were used to develop rules for the linguistic models and test the resulting model. Results show a 0.76 correlation between the model and the actual split size for test runs. Performance loss is attributed to poor representation of the system by the training data and the large number of search parameters.

Summary

Empirical screen models use data collected during typical operating conditions. Regression methods are used to relate input variables to output variables resulting in a screen specific model. Capacity models were found to be easily understood, however, the performance of capacity models are limited [15]. Models that related screening variables to the screens efficiency curve were reviewed and efficiency models by King [26] and Karra [24] were presented. Finally, several models were presented that use advanced statistical regression methods to relate input variables to output variables. The advantage of empirical models is that they are relatively simple and are easily applied to screen performance prediction and operation. Furthermore, if no applicable model exists for the screen of interest, empirical methods can be used to develop a new model. However, in all cases, the resulting

empirical screen models are limited to the quality and range of data collected. The first two groups of empirical models rely on the feed particle size distribution data as the main component for predicting screen performance. Since time series feed particle size distribution data is not available, the sole empirical method that remains is the advanced statistical regression approach which can be used with available historical time series data to build an oil sand screen model.

2.4.2 Probabilistic Models

Probability models look at screening phenomena as a series of probabilistic events, where each time a particle interacts with the screen it will have a certain probability of passing through. The probability of a particle passing through the screen is dependent on both design and operating screen variables. Probability models can also be defined as the probability of a particle not passing through the screen, which forms an overflow partition curve. An overflow partition curve is simply the percentage of particles in each size bin that will probabilistically pass to overflow. Figure 2.4 shows a typical overflow partition curve for a vibrating screen.

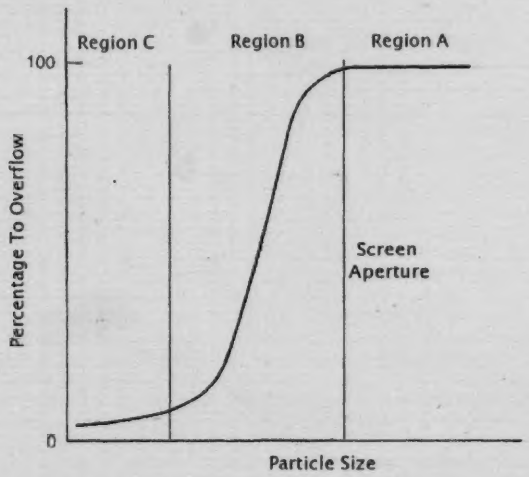

Figure 2.4: A typical overflow partition curve for a vibrating screen.

The typical overflow partition curve has three regions [36]. The first region, Region A, is simple to model since all particles are larger than the screen aperture and pass to overflow. Region B contains particles that are within 75 percent of the aperture diameter and are the focus of most probability models. Region C contains particles that are less than 25 percent of the aperture diameter. Models for Region C are rarely developed as the particles in this region are assumed to easily pass through the screen apertures.

The area of probabilistic modelling has been well researched [19, 32, 9, 59, 53]. Two well-known

probability models, the Gaudin model and the Whiten model, are outlined below.

Gaudin Model

Gaudin first introduced his probability model in 1939 [19], noting that the probability of a round particle passing through a square aperture, $p(x)$, could be described by:

$$p(x) = \frac{(h-x)^2}{(h+d)^2} \approx \left[1 - \frac{x}{h}\right]^2 \text{ for d << h} \tag{2.11}$$

where h is the aperture size, d is the wire diameter, and x is the particle size. Each time the particle hits the screen, the probability of it passing through increases. The probability of a particle reaching the end of the screen without passing through a screen aperture would be $(1 - p(x))^N$, where N is the number of times the particle hits the screen. Figure 2.5 is a plot of Equation 2.11 and shows that the resulting overflow partition curve is similar to that of Region B in Figure 2.4.

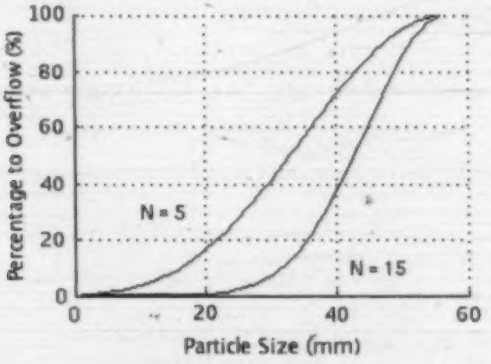

Figure 2.5: Graph of Gaudin's overflow partition curve for N = 5 and N = 15. The aperture size is 55 mm. Note the similar shape to that of Region B in Figure 2.4.

The model proposed by Gaudin is simple and has been used as a basis for several probability models [36, 59, 53]. Its simplicity, however, does not extend to particle-particle interaction and fails to model complex screening behaviour when the screen bed is more than one particle high. In the next section we show how Whiten et al. have extended the Gaudin model to real world screening.

Whiten Model

Whiten's model [59, 61, 60] is based on Gaudin's work on screening theory. The resulting model put forward by Whiten is

$$E(x) = exp\left[-N f_o \left(1 - \frac{x}{h}\right)^k\right] \tag{2.12}$$

where N is the number of passage attempts along the screen, f_o is the percentage of open area, x is the particle size, h is the aperture size, and k is a fitting parameter typically set at 2. All variables in

Equation 2.12 are easily measured except for N, the number of passage attempts. There is currently no simple means to measure the number of times each particle hits the screen. An alternative is to indirectly measure the value through regression of N to collected data such as the feed tonnage and the material used in the screening media.

Whiten also developed a model to estimate the passage of fines to overflow for Region C in Figure 2.4. The passage rate is estimated by a fines factor that is calculated using $A_s \times SF$, where A_s is the total surface area of the particles and SF is a fitting parameter. The fitting parameter is a function of the percentage of fines in the feed and feed rate of the fines material. To determine the total surface area of the particles, Whiten proposes Equation 2.13.

$$A_s \propto \sum_{i=1}^{n} \left[\frac{VOL_i}{\frac{d_i + d_{i+1}}{2}} \right] \tag{2.13}$$

where d_i is the mean diameter of particles in size bin i. Hess notes that Whiten's model requires no special sampling techniques which increases its applicability in modelling existing screening systems. The simplicity of this model is in its implementation and has been used successfully in several pilot and online plants [61] [59].

Summary

Probability models use probability theory to model the screening phenomena, and compared to empirical models, are based on a greater understanding of screening physics. Probability models are popular for aggregate industrial screening and have been successfully used in modelling real screening applications [59, 28, 61]. However, the main screening variable used in probability models is the feed particle size distribution, which is used to estimate the overflow partition curve. Unfortunately time series feed particle size distribution data not available at Syncrude's North Mine. Therefore probability models cannot be considered for modelling oil sand static screens at this time.

2.4.3 Kinetic Models

The last category of screen models is kinetic models. Kinetic models are the most studied of all screen models as they have been shown to provide the most accurate modelling of test screens. Kinetic models measure the passage rate of material along the screen's length. The material passage rate depends largely on the size of the particles. Very large particles will never pass through and very small particles will pass through the first time they see the screen. Figure 2.6 shows a typical passage rate curve for a vibrating screen.

The passage rate is a function of screening variables such as the feed particle size distribution, the feed tonnage and the aperture size. Referring to Figure 2.6, typically the first region is considered very short and is often hard to see [54]. The second region, referred to as the crowded region, occurs when sufficient material is placed on the screen to form a bed. In this region, stratification of the bed occurs when smaller particles move through spaces created by larger particles. The third region,

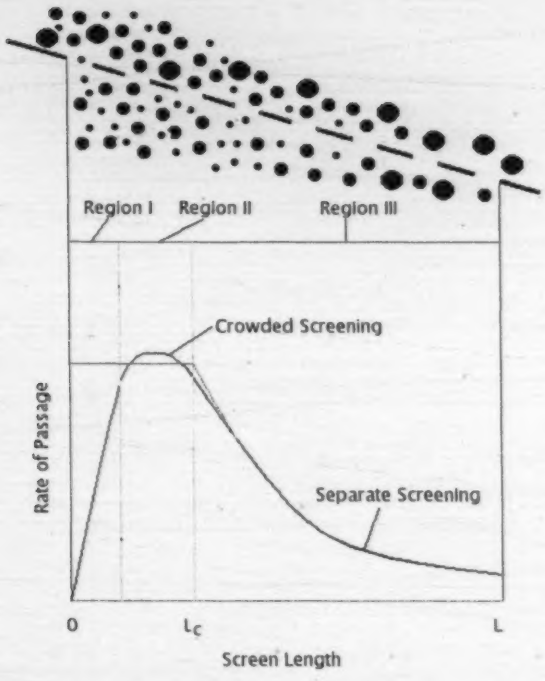

Figure 2.6: The passage rate as a function of the position along the screen. Also marked are the crowded and separated regions on the screen.

referred to as the separate region, occurs when a single layer of material exists on the screen and little interaction between particles occurs. The transition point, L_c, is located where the crowded region ends and the separate region starts.

Table 2.5 gives a quick summary of definitions for common variables used in the following overview of kinetic screening theory.

Variable	Definition
$w_x(l)$	Flow rate of feed per unit width in size interval x at position l on the screen.
$W(l)$	Total flow rate of feed per unit width at position l on the screen.
$m_x(l)$	Mass of feed per unit width in size interval x at position l on the screen.

Table 2.5: List of kinetic variables used in the overview of kinetic screening theory.

Kinetic modelling assumes steady-state screening conditions, dry screening material, and the feed entering the screen is assumed to be perfectly mixed. A simple kinetic model and the Ferrara kinetic model are now presented.

Simple Kinetic Model

A simple kinetic model is presented by King [26]. Equation 2.14 shows the general passage rate equation, where the change in flow rate of material at point l along the screen, $\partial w_x(l)/\partial l$, is equal to the amount of material that passes through the screen, or the rate of transmission r_x.

$$\frac{\partial w_x(l)}{\partial l} = -r_x \tag{2.14}$$

The rate of transmission in the crowded region in size bin x, r_x, is $k_x m_x(0)$, where k_x is the kinetic constant for the crowded region. The kinetic constant captures how quickly material in each size bin passes through the screen. The rate of transmission is then a function of the kinetic constant and the amount of material on the screen, m_x. The rate of transmission is assumed constant across the crowded region and so the initial value of m_x can be used. The constant rate of transmission is due to particles from the bed presenting themselves to the screen at the same rate particles on the screen surface pass through. This creates a steady flow of material through the screen in the crowded region. Substituting $r_x = k_x m_x(0)$ into Equation 2.14 and integrating the length of the crowded region from 0 to L_c, the passage rate at position L_c along the screen is written as:

$$w_x(L_c) = w_x(0) - k_x m_x(0) L_c \tag{2.15}$$

Using $w_x(l) = m_x(l)u$, where u is the average speed at which the particles are moving, Equation 2.15 can be rewritten as:

$$w_x(L_c) = m_x(0)(u - k_x L_c) \tag{2.16}$$

Assuming that the rate of transmission in the crowded region is constant then k_x is not dependent on l. To determine L_c, King assumes that the transition point occurs when the smallest size fraction of the material is no longer present on the screen. Once the position of L_c is known, the remainder of the values of k_x can be determined through regression to collected data.

The separate region is easier to model as it assumes that there are no interactions between particles. The separate region is also described by Equation 2.14, where $r_x = s_x m_x(l)$ and s_x is the kinetic constant for the separate region.

$$\frac{\partial w_x(l)}{\partial l} = -s_x m_x(l) = -\frac{s_x}{u} w_x(l) \tag{2.17}$$

Integrating Equation 2.17 from L_c to l gives:

$$w_x(l) = w_x(L_c) exp\left(-\frac{s_x}{u}(l - L_c)\right) \tag{2.18}$$

Equation 2.18 gives the complete mass flow rate of material in the separate region of screening. Combining Equation 2.15 and 2.18 we get the passage rate at any point along the screen.

$$w_x(l) = [w_x(0) - k_x m_x(0) L_c] exp\left(-\frac{s_x}{u}(l - L_c)\right) \tag{2.19}$$

where s_x can be approximated by probability theory for single particle screen interaction.

King determines the partition factor, C_x, also known as the overflow partition curve, from equation 2.19, where $C_x = w_x(L)/w_x(0)$.

$$C_x = \left[1 - \frac{r_x L_c}{w_x(0)}\right] exp\left(-\frac{s_x}{u}(l - L_c)\right) \tag{2.20}$$

The simple kinetic model provides a quick introduction to kinetic modelling theory. However, the assumption that the rate of transmission in the crowded region is not dependent on the distance along the screen often fails in reality. The next model, proposed by Ferrara et al., includes the change in rate of transmission along the screen to produce a more accurate model of screen passage rates in the crowded region.

Ferrara Model

Ferrara and Preti [17] also proposed a zero-order process for the crowded region and a first-order process for the separate region. Unlike the simple kinetic model, where the load M on the screen is assumed constant throughout the crowded region, the Ferrara model assumes that the load on the screen changes with position l. The zero-order process used to represent the crowded portion of the screen is written as:

$$\frac{\partial w_x(l)}{\partial l} = -k_x^f \cdot f_x(l) \tag{2.21}$$

where k_x^f is the rate constant for crowded screening conditions, and $f_x(l)$ is the fraction of material in size bin x at position l. Note that the kinetic constant introduced in the simple kinetic model is not equivalent to the rate constant presented here. The material density variable can also be rewritten as:

$$f_x(l) = \frac{w_x(l)}{W_x} = \frac{w_x(l)}{\int_0^\infty w_X(l) dX} \tag{2.22}$$

where X can be any of the other bin sizes including x. Ferrara defines a new variable $\chi = ln(E_X(l))/ln(E_x(l))$, where $E_x(l)$ is the fraction of material remaining on the screen in size bin x at position l. The variable $E_x(l)$ can also be written as $w_x(l)/w_x(0)$. Rearranging Equation 2.22 and substituting in $w_x(l) = w_x(0)E_x(l)$, $w_X(l) = w_X(0)E_X(l)$, and $E_X(l) = E_x(l)^\chi$ gives:

$$\int_0^\infty w_X(0) dX E_x(l)^{\chi - 1} \frac{\partial E_x(l)}{\partial l} w_x(0) = -k_x^f \tag{2.23}$$

Now Equation 2.23 is in a form that can be used to investigate k_x^f, which is the unknown, and how it relates to operating variables such as the aperture size. Ferrara et al. proposed that k_x^f should be represented as:

$$k_x^f = k_{50}^f 2^\sigma \left[1 - \frac{x}{h}\right]^\sigma \tag{2.24}$$

where h is the size of the screen aperture, k_{50}^f is the kinetic constant for when $x/h = 0.5$, and σ is a fitting factor.

The form of the model for the separate region is similar to that of the simple model presented earlier and is written as:

$$\frac{\partial w_x(l)}{\partial l} = -s_x^f w_i(l) \tag{2.25}$$

where s_x^f is the kinetic constant for size bin x. Ferrara et al. propose that s_x^f can be modelled as $s_x^f = nln[1 - p_x]$, where n is the number of times the particle sees the screen and p_x is the probability of passage of a particle of size x. Using Gaudin's expression for the probability of passage of a particle written as:

$$p(x) = \left(\frac{h-x}{h+d}\right)^\sigma = f_0^{\sigma/2}\left(1 - \frac{x}{h}\right)^\sigma \tag{2.26}$$

where σ is typically set to 2, f_o is the fraction of open area, and h is the diameter of the aperture. Integrating Equation 2.26 from L_c to L gives:

$$E_x(L) = exp\left(-nln\left[1 - f_0^{\sigma/2}\left(1 - \frac{x}{h}\right)^\sigma\right] L\right) \tag{2.27}$$

By fitting Equation 2.27 to collected data, n can be determined for a specific set of screening conditions. The kinetic model proposed by Ferrara et al. is more complicated and for practical use requires the rate equation for the crowded region to be solved using finite difference approximation [26].

Feed Bed Stratification

The stratification rate is the speed at which particles pass through the material bed to arrive at the screen. Figure 2.7 shows the stratification of particles within a bed of material.

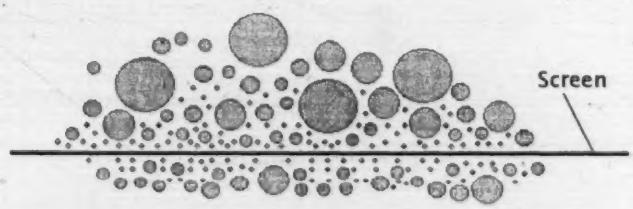

Figure 2.7: The stratification of particles within a bed of material.

The rate of stratification depends on the relative particle sizes in the feed, the particle density curve and the screen motion [55]. Soldinger [49] also notes the influence of bed thickness on

the screening process. Stratification is known to aid screening, enabling particles to quickly move through material to the screen below. Stratification is found to increase with a greater breadth in feed particle size distribution allowing smaller particles to fall between larger particles. The level of stratification will decrease when the average particle size decreases. Finally, wet conditions hinder stratification as smaller particles are more likely to stick to larger particles [49]. It was noted that particle density and particle shape have less of an influence on the rate of stratification as compared to the particle size distribution [33].

The rate of stratification is typically modelled as a first order process [55, 25]. Referring to Soldinger's work on stratification [49], if the amount of material in the bottom layer is S, and the remaining amount of material in the screen bed is $1 - S$, then the rate of stratification of material is written as:

$$\dot{S} = c(1 - S) \tag{2.28}$$

where c represents the rate of stratification. Soldinger observed that the value of c varied along the length of the screen. Integrating Equation 2.28 gives:

$$S = 1 - (1 - S_0)exp(-ct) \tag{2.29}$$

Modelling particle stratification within the material provides important additional information of the particle behaviour on the screen.

Summary

Kinetic models use passage rate information to model screen performance. Knowing the passage rate along the screen at different lengths, the percentage of overflow material for each particle size bin can be determined. In both the simple kinetic screen model and the Ferrara kinetic screen model the screen was divided into a crowded region and a separate region, which were modelled separately. The simple kinetic model was shown to be easily derived, however the kinetic screen model proposed by Ferrara et al. was shown to more realistically model the crowded region. Additional kinetic screen models are available that use similar assumptions and result in similar model forms [52, 16, 63, 6]. Overall, kinetic models have performed well in controlled experiments [22] [18] [50].

To implement a kinetic screen model, the underflow tonnage is required at regular intervals along the screen. This is easily accomplished in a pilot plant situation, but can be a time consuming and costly undertaking for an operational plant. Consequently, the kinetic model is not a feasible option for the analysis of an operational plant screening system such as the one investigated in this work.

2.4.4 Evaluation of Traditional Screen Models

In Sections 2.4.1 to 2.4.3 three categories of screen models were covered and it was found that only the category of empirical models that use advanced statistical regression methods is applicable to oil

sands screen modelling at this time. However, of the traditional screening variables listed in Section 2.3, time series data is only available for the feed tonnage variable. A representative sample of data points is plotted to determine whether a simple relationship does exist between the feed tonnage and the overflow tonnage. If a high percentage of the overflow tonnage variance is explained by the feed tonnage alone, then the remaining traditional variables can be considered to play a relatively minor role in oil sand screen performance.

Figures 2.8 and 2.9 plot the overflow tonnage versus the feed tonnage for two years of data collected during summer months (May to August) and two years of data collected during the winter months (December to March) respectively. Both plots show that the feed tonnage is in general a poor predictor of the overflow tonnage. This result motivates the need for further understanding of the overflow tonnage variance as it relates to other screening variables. Using statistical regression methods, pertinent process variables, from Syncrude's historical databases, can be identified and used to improve prediction of overflow tonnage variance.

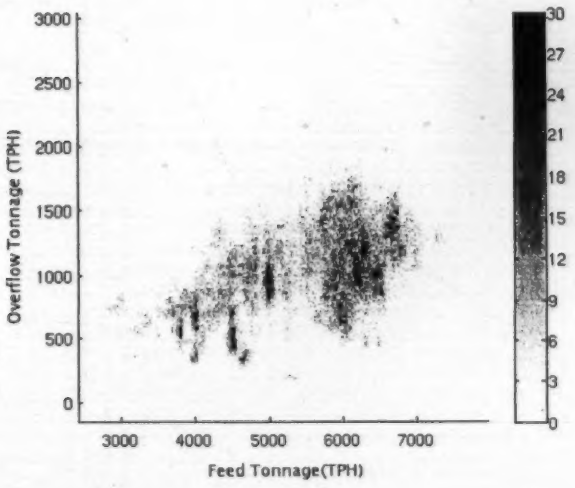

Figure 2.8: A gray coded scatter plot of overflow tonnage versus feed tonnage during summer months (May to August) for 2005 and 2006 for the screening system in Figure 1.2. The varying levels of gray represent the local data point density ranging from zero (white) to 30 (black).

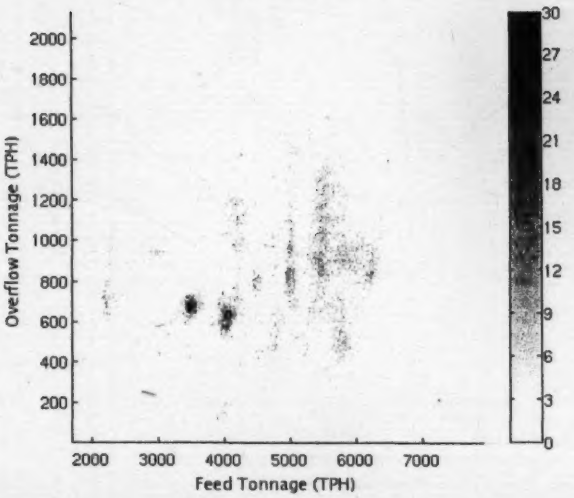

Figure 2.9: A gray coded scatter plot of overflow tonnage versus feed tonnage during winter months (December to March) for 2005 and 2006 for the screening system in Figure 1.2. The varying levels of gray represent the local data point density ranging from zero (white) to 30 (black).

2.5 Summary

It was shown in this chapter that traditional screen models depend heavily on the feed particle size distribution. However, time series feed particle size distribution is currently not available for Syncrude's North Mine. Before undertaking the extensive work needed to collect data for traditional screening variables, plant variables are considered such as the screening system's water usage and oil sand geological properties. Taking an empirical approach, statistical regression methods, introduced in Chapter 3, are used to build a model relating water and geological variables to screen performance.

Chapter 3

Regression Techniques

The previous chapter introduced traditional screening variables and screen models. Unfortunately, traditional screen models could not be evaluated on Syncrude's static screens due to the limited available historical time series data. Consequently, an empirical statistical regression screen modelling approach is adopted. In this chapter, we outline the general approach taken for the development of a screen model using historical time series data and present in detail four statistical regression methods.

3.1 Introduction

Statistical regression methods have been applied to many different research areas in order to gain insight into complicated systems. They enable researchers to isolate potential relationships between predictor variables and response variables without requiring in-depth knowledge of the underlying system. In this chapter we introduce a quick overview of the experimental process as applied to the analysis of historical data. We overview four well-known statistical regression methods: multiple linear regression, response surface regression, principal component regression, and partial least squares regression. Finally, we discuss tools used to evaluate and extract conclusions from regression results.

3.2 Outline of Regression Model-Building Process

In this work, a general approach is taken to identifying significant variables and the development of a regression model. A typical strategy for building a regression model is presented in Figure 3.1.

The first stage provides the problem formulation that addresses the goals of the experiment. Following that, the data is explored to remove obvious errors, extreme outliers, and provide summary statistics. Summary statistics include the mean, variance and range of the data. The second stage is to screen the input variables using linear models to identify which are the most significant. During the process of identifying significant variables, an explanatory model is developed to provide insight into the general oil sand screening process. In the third step, prediction models are developed with more

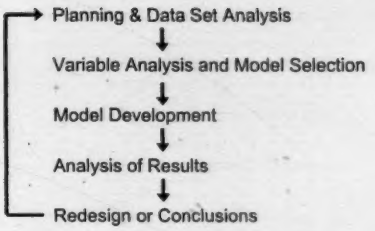

Figure 3.1: Regression model-building process [37].

advanced regression methods to increase the predictive power of the model. Prediction regression methods typically include interaction terms, polynomial terms, and nonlinear functions for each predictor variable. Four regression methods used in this work are outlined in the following section. The fourth stage analyzes the goodness of fit of the developed model. Common forms of result analysis are coefficient analysis with 95% confidence intreval, RMS error, f-statistics, t-statistics, variance inflation factors, and the coefficient of determination. The last stage considers the results and evaluates whether satisfactory results have been achieved or if the process should be repeated.

3.3 Statistical Regression Methods

In this section we overview multiple linear regression, response surface regression, principal component regression, and partial least squares regression. In the presentation of each regression method, M is the number of variables, including the bias term, and N is the number of data points. The mean of the response matrix is denoted by $\bar{Y}$ and the standard deviation of the response matrix is denoted by a $\hat{Y}$. Data used to train a screen model is denoted by a $train$ subscript and data used to test a screen model is denoted by a $test$ subscript.

3.3.1 Multiple Linear Regression

Multiple linear regression (MLR) is a statistical regression method for determining linear relationships between multiple predictor variables and a single response variable. MLR is widely used for its simplicity and the ease with which the resulting coefficients can be interpreted.

The functional form of MLR is given in Equation 3.1, where Y is a Nx1 matrix of response variable data points, B is a Mx1 matrix of system coefficients, X is a NxM matrix of predictor variable data points, and ϵ is an error term that represents the remaining system variance not captured by the model. Typically, the model coefficients are determined from a training data set using least squares estimation.

$$Y_{train} = X_{train} \cdot B + \epsilon \tag{3.1}$$

Once the coefficients are determined from the training data set, testing of the model is easily carried out. Each data point of the test set, X_{test}, is multiplied by the coefficients and the resulting values are linearly summed to give a predicted value for the response variable. The functional form is provided in Equation 3.2.

$$Y_{test} = X_{test} \cdot B \tag{3.2}$$

MLR assumes that the predictor variables are measured with little error and that the relationships between the two sets are approximately linear [35]. MLR also assumes that the predictor variables are independent since MLR is simply a series of univariate regressions. A disadvantage of MLR is that if predictor variables are not independent, problems of multicollinearity arise that can limit performance [30]. Additional assumptions include non-stochasticity in the predictor variables and statistical independence and normal distribution, with zero mean and constant standard deviation, of model errors. Preprocessing of data using centring or scaling techniques is not necessary due to the linear properties of MLR.

To summarize, MLR is a simple yet powerful method for quick evaluation of data from complex systems. Despite a number of limitations, MLR's strength is in its ability to provide information about the system by analyzing the resulting coefficients. Preliminary relationships can be identified between response variables and significant predictor variables, which could not easily be accomplished using more advanced regression methods.

3.3.2 Response Surface Regression

Response surface regression (RSR) is an extension of MLR that includes interaction terms and higher order terms. In this work we limit the higher order terms to at most degree two. An example of the response surface regression form for two predictor variables is given in Equation 3.3.

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_1^2 + b_4x_2^2 + b_5x_1x_2 + \epsilon \tag{3.3}$$

A general form of the RSR model is $Y = XB + \epsilon$, where Y is a matrix of response variable data points, B is a matrix of system coefficients, X is a matrix of predictor variable data points including all interaction and quadratic terms, and ϵ is an error term that represents the remaining system variance not captured by the model. For RSR, both the predictor variables and the response variable are centred and scaled. As in MLR, the coefficients are determined using least squares estimation.

Prediction using a test data set is accomplished by subtracting the mean of the training data set, X_{train}, from the test data set, X_{test}, and then multiplying by the coefficients determined from least squares estimation. The result is then multiplied by the standard deviation and added to the mean of the original response matrix, as shown in Equation 3.4.

$$Y_{test} = \left[\frac{X_{test} - \bar{X}_{train}}{\hat{X}_{train}}\right] \cdot w \cdot B \cdot \hat{Y}_{train} + \bar{Y}_{train} \tag{3.4}$$

Response surface regression enables modelling of relationships caused by the interaction of variables, but suffers from overfitting if proper care is not taken. Response surface regression is often used in applied research [8].

3.3.3 Principal Component Regression

Principal component regression (PCR) addresses the problem of multicollinearity between predictor variables. In PCR, principal component analysis (PCA) is used to decompose the predictor matrix X into a series of orthogonal components. Regression methods then use a subset of the resulting components to develop a model of the system.

The system takes the linear form $Y = XB + \epsilon$, where Y is a matrix of response variable data points, B is a matrix of system coefficients, X is a matrix of predictor variable data points, and ϵ is an error term that represents the remaining system variance not captured by the model. PCA generates a set of eigenvectors, w, and eigenvalues of $X^T X$. Both Y and X are first centred by subtracting their respective means and scaled by dividing by their respective standard deviations. Each eigenvector accounts for a portion of the system variance. The magnitude of the variance is measured by the magnitude of the eigenvalue. The PCR components are then determined by multiplying the eigenvectors by the original predictor matrix. Typically, the number of components chosen account for 80 to 90 percent the total system variance, where the remaining components are assumed to represent noise [29]. Figure 3.2 shows a simple example of principal components in a two-dimensional data set.

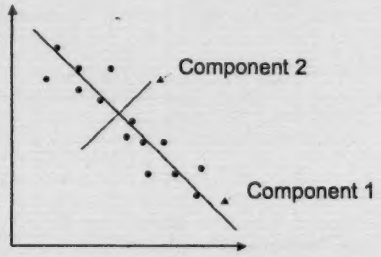

Figure 3.2: Example of principal components in a two-dimensional data set.

The resulting subset of eigenvectors, w, are then used to decompose the original data set into orthogonal components by multiplying each data point by the set of eigenvectors. The new predictor matrix is now in the form Xw. Once the new predictor matrix is formed, then MLR is performed, using Equation 3.5, to determine coefficients, B, that relate the response variable to the new predictor matrix.

$$Y_{train} = \left[\frac{X_{train} - \bar{X}_{train}}{\hat{X}_{train}}\right] \cdot w \cdot B \cdot \hat{Y}_{train} + \bar{Y}_{train} \tag{3.5}$$

Prediction using a test data set is accomplished by subtracting the mean of the training data set, X_{train}, from the test data set, X_{test}, and then multiplying by the coefficients determined from MLR. The result is then multiplied by the standard deviation and added to the mean of the original response matrix, as shown in Equation 3.6.

$$Y_{test} = \left[\frac{X_{test} - \bar{X}_{train}}{\hat{X}_{train}}\right] \cdot w \cdot B \cdot \hat{Y}_{train} + \bar{Y}_{train} \tag{3.6}$$

It was discussed earlier that MLR can behave poorly in data sets with highly correlated variables. The advantage of using PCA prior to using MLR, is that the new data set satisfies the assumption of orthogonality. A second advantage of PCR is the inherent data reduction gained from using only a subset of the total number of components. For example, in Figure 3.2, the data can be represented using simply the first component allowing for a small amount of error in the final results. One disadvantage of PCR is that PCA only considers the variance of the predictor matrix. Consequently, if the components with the lower percentage of explained variance are the more important variables then they can be removed inadvertently when selecting the subset of principal components. This results in the selection of a sub-optimal set of components. Therefore a better method, presented in the following section, is to evaluate components not based solely on the variance in the predictor variables, but also on the variance in the response variable.

3.3.4 Partial Least Squares Regression

Partial least squares regression (PLSR) identifies latent vectors that maximize the explained variance between the predictor variables and the response variable [29]. Unlike PCR, where the top k components may not be the optimal set of principal components, the top k latent vectors in PLSR are optimal for the system as they are determined by maximizing the explained variance of both the predictor and the response variables.

As before, the system is assumed to take the form $Y = XB + \epsilon$, where where Y is a matrix of response variable data points, B is a matrix of system coefficients, X is a matrix of predictor variable data points, and ϵ is an error term that represents the remaining system variance not captured by the model. Once again the predictor and response matrices are centred by subtracting the mean and scaled by dividing by the standard deviation.

A commonly used method for identifying latent vectors in a system is Nonlinear Iterative PLS (NIPALS) [29]. NIPALS builds a weight matrix, w, that de-correlates X. The general algorithm for NIPALS is reproduced in Table 3.1.

Referring to Table 3.1, NIPALS finds the latent vectors w_h and then subtracts the resulting explained variance from the predictor and response matrices. This is repeated until a preset percentage of the system variance is represented by the identified latent vectors. At this point we have a set

1 · $X_1 = X$, $Y_1 = Y$, $C_1 = I$
2 · For h = 1,....,N
3 · $w_h = \text{norm}(C_h X_h' Y_h)$
4 · $t_h = X_h w_h$
5 · $Y_{h+1} = Y_h - (t_h t_h' Y_h)/(t_h' t_h)$
6 · $C_{h+1} = C_h - (w_h t_h' X_h)/(t_h' t_h)$
7 · $X_{h+1} = X_h - (t_h t_h' X_h)/(t_h' t_h)$
8 · End For

Table 3.1: NIPALS pseudo-algorithm [29], where N is the number of desired latent vectors.

of weight vectors, or latent vectors, that are used to decompose the predictor matrix as shown in Equation 3.7.

$$Y_{train} = \left[\frac{X_{train} - \bar{X}_{train}}{\hat{X}_{train}}\right] \cdot w \cdot B \cdot \hat{Y}_{train} + \bar{Y}_{train} \tag{3.7}$$

Similar to PCR, the new decorrelated predictor matrix is then fit to the response matrix using MLR or a similar regression method. Prediction is carried out in the same manner as PCR. The result is then multiplied by the standard deviation and added to the mean of the original response matrix, as shown in Equation 3.8.

$$Y_{test} = \left[\frac{X_{test} - \bar{X}_{train}}{\hat{X}_{train}}\right] \cdot w \cdot B \cdot \hat{Y}_{train} + \bar{Y}_{train} \tag{3.8}$$

An example is often helpful in understanding the difference between PCR and PLSR. In Table 3.3 we have a response variable and two predictor variables with a series of values. Table 3.4 gives the centred and scaled values of the original data. As we can see from a quick glance, to predict the response variable, only the second predictor variable is needed.

Y	X_1	X_2
1	20.2	1
2	13.4	2
3	21.2	3
4	34.5	4
5	6.7	5

Figure 3.3: Example values for a response vector, Y, and two predictor variables, X_1 and X_2.

Y	X_1	X_2
-1.26	0.096	-1.26
-0.63	-0.56	-0.632
0	0.19	0
0.63	1.48	0.632
1.26	-1.21	1.26

Figure 3.4: Values of the response and predictor variables after centring and scaling.

Both PCR and PLSR are run on the scaled and centred data in Table 3.4. Figure 3.5 plots the data with the PCR components and Figure 3.6 plots the data with the PLSR latent vectors.

We can see that PCR ignores the values of the response variable and identifies the maximum variance in the predictor variables only. From Figure 3.6 we can see that PLSR takes into consideration the variance of the response variable and, while maintaining orthogonality in the latent vectors, rotates the components accordingly. Table 3.2 gives the results of applying multiple linear regression to the first component of both PCR and PLSR and the response vector.

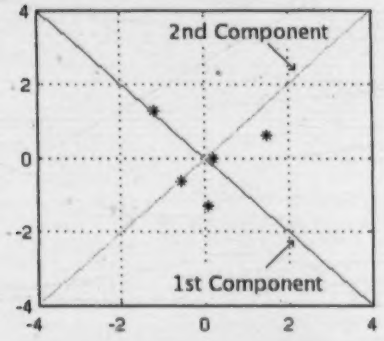

Figure 3.5: Plot of centred and scaled example data with overlaid PCR components.

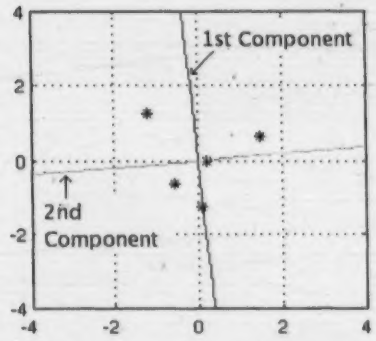

Figure 3.6: Plot of centred and scaled example data with overlaid PLSR latent vectors.

Y	Y_{PCR}	Y_{PLSR}
1	1.92	1.01
2	2.94	1.99
3	2.85	2.97
4	2.33	3.78
5	4.95	5.14
Error	0.75±0.66	0.10±0.08

Table 3.2: Final results using MLR on only the first resulting component from PCR and PLSR respectively.

From Table 3.2, we can see that PLSR provides a stronger first component, or latent vector, because it takes into account the variance in the response vector.

3.4 Presentation of Results from Regression Analysis

In this section we present several typical regression performance measures that are used for the evaluation of experimental results. First, an example is provided to aid in explaining the regression performance measures. Table 3.3 lists example data for five predictor variables and one response variable.

A multiple linear regression model is built relating the predictor variables to the response variable. The resulting performance measures are presented in Table 3.4.

In Table 3.4, in the top section under the source heading, *model* refers to the regression model, *total* refers to the system under investigation, and *residual* refers to the difference, or error, between the model and the system. Table 3.5 summarizes the variable symbols used in the following discussion of regression performance measures.

Regression models are commonly evaluated using the residual root mean squared error (RMS error). The residual RMS error provides an evaluation of the average error between the model and measured value. The residual RMS error is calculated using Equation 3.9.

Response	Var. 1	Var. 2	Var. 3	Var. 4	Var. 5
5.00	-1.56	1.45	169.5	15.69	3.45
9.00	-3.56	3.45	200.1	16.23	2.99
3.33	-0.56	0.99	212.1	15.45	3.33
1.45	-4.50	1.23	145.2	14.23	4.42
12.7	-3.69	4.56	123.3	13.69	2.58
20.4	-0.23	1.56	134.9	12.12	3.45
15.4	-1.34	1.65	111.2	18.23	6.23
1.23	-6.75	2.34	178.4	19.23	4.56
23.0	-3.45	4.34	194.4	16.56	3.24
5.69	-1.23	2.45	138.2	15.99	1.23

Table 3.3: Example data set for one response variable and five predictor variables.

Response Variable: Overflow Tonnage

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F-Value	Prob > F
Model	5	428.61	85.72	2.78	0.1717
Residual	4	123.30	30.83		
Total	9	551.91			

RMS Error	5.55
R^2	0.777
adj-R^2	0.497

Parameter Estimates

Variable	DF	Reg. Coef.	Standard Error	T for H0	Prob > |T|
Constant	1	1.91	16.98	0.11	0.916
Var 1.	1	2.82	1.14	2.47	0.069
Var 2.	1	6.16	1.84	3.35	0.029
Var 3.	1	0.012	0.06	0.20	0.855
Var 4.	1	-0.996	1.12	-0.89	0.423
Var 5.	1	4.06	1.91	2.13	0.101

Table 3.4: Resulting performance measures from multiple linear regression performed on the example data set in Table 3.3.

Variable	Symbol	Variable Definition
Sample Overflow Tonnage	Y_{sample}	
Model Overflow Tonnage	Y_{model}	
Residual Values	RES	$RES = Y_{model} - Y_{sample}$
Total Sum of Squares	SS_{total}	$(Y_{sample} - mean(Y_{sample}))^2$
Residual Sum of Squares	$SS_{residual}$	$(RES - mean(RES))^2$
Model Sum of Squares	SS_{model}	$SS_{model} = SS_{total} - SS_{residual}$
Total Degrees of Freedom	DF_{total}	DF_{total} = number of data points minus 1.
Model Degrees of Freedom	DF_{model}	DF_{model} = number of predictor variables.
Residual Degrees of Freedom	$DF_{residual}$	$DF_{residual} = DF_{total} - DF_{model}$

Table 3.5: Variables used in the discussion of regression performance measures.

$$\text{Residual RMS Error} = \sqrt{\frac{SS_{residual}}{DF_{residual}}} \tag{3.9}$$

The coefficient of determination, R^2, is the proportion of variability in the response variable that is explained by the model. The coefficient of determination is calculated as:

$$R^2 = \frac{SS_{residual}}{SS_{total}} \tag{3.10}$$

where a value of one indicates that the model has perfect predictability and a value of zero implies that the model has none. Unfortunately, R^2 can be increased by adding new variables even though the new variables add little predictive capability to the model. The adjusted coefficient of determination, $adj\text{-}R^2$, accounts for variables that provide little or no new predictive capability. The adj-R^2 is calculated using Equation 3.11.

$$adj\text{-}R^2 = 1 - (1 - R^2)\frac{DF_{total}}{DF_{residual}} \tag{3.11}$$

In Table 3.4, the standard error (SE), or standard deviation, of the coefficient value is used for hypothesis testing and constructing a confidence interval. The standard error is calculated as:

$$SE = \sqrt{diag\,(R^{-1} \cdot (R^{-1})^T \cdot MSE)} \tag{3.12}$$

where R is calculated from a QR decomposition of X and MSE is the residual mean squared error. The confidence interval is constructed by multiplying the SE by k, where k is a constant related to the level of confidence desired. For a 95 percent confidence limit, the value of k is 1.96.

The t-statistic tests the hypothesis that the regression coefficient for the population is zero in the presence of the remaining predictor variables. It is calculated as:

$$\text{t-statistic} = \frac{\text{regression coefficient} - \text{hypothesized value}}{SE} \tag{3.13}$$

where the hypothesized value is set to zero in order to evaluate if the regression coefficient is zero in the presence of the remaining variables. The t-statistic is then mapped to a normal distribution to determine the P-value of the coefficient. The P-value is the observed significance levels for the t-statistic. The P-value is used to identify variables that could potentially be removed without significantly reducing the model's predictive power. Insignificant variables should be removed one at a time since a variable that is insignificant in the presence of a set of variables, may become significant when a variable from that set is removed. If a P-value has an insignificant value ($> 5\%$) then that variable can be considered for removal.

The F-value is a statistic used to evaluate the model as a whole. The F-value is calculated using Equation 3.14.

$$F = \frac{MSE_{model}}{MSE_{total} - MSE_{model}} \tag{3.14}$$

where the mean squared error (MSE) is calculated by dividing the sum of squares value by the number of degrees of freedom for the *model* and the *residual* respectively. This value is used to determine if the null hypothesis, indicating that the model has no predictive capability, should be rejected. The model is rejected if the F-value is less than zero, when $SS_{total} < SS_{residual}$.

Not shown in Table 3.4 are variance inflation factors (VIF) that evaluate multicollinearity between variables [38]. If the predictor variable is uncorrelated with the remaining variables then the value of the variance inflation factor is equal to one. A larger value indicates a higher level of redundancy of predictor variables. Typically, a predictor variable having a variance inflation factor greater than 10 should be considered for removal. The VIF is calculated using Equation 3.15.

$$VIF_i = \frac{1}{1 - R_i^2} \tag{3.15}$$

where i is the variable of interest and R_i^2 is the unadjusted R^2 when you regress X_i against the remaining predictor variables in the model. Once highly correlated variables are identified and removed, a stronger assumption of independence can be made for MLR.

The regression models developed in this work are evaluated using the RMS error and the adj-R^2 value. Further significance testing is carried out using t-statistics and VIF. The RMS error provides a measure of the average error of the model and the adj-R^2 value provides a measure of the system variance explained by the model. Together, they provide a measure of model performance and allow the evaluation and ranking of different regression models created using the same data set.

3.5 Summary

In this chapter, a general approach for the development of a screen model was introduced along with four statistical regression methods: multiple linear regression, response surface regression, principal component regression, and partial least squares regression. The next chapter will outline steps taken to collect data, identify significant variables, and develop explanatory and predictive screen models.

Chapter 4

Model Development

In the previous chapter, we outlined the general approach to regression data analysis and overviewed four common linear regression methods. In this chapter we collect and clean raw water and geological data and identify significant variables for summer and winter screening conditions using forward stepwise multiple linear regression. We show that a single screen model is sufficient for summer screening conditions for both Train 1 and Train 2, however, separate models are needed for Train 1 and Train 2 in winter screening conditions. Candidate prediction models are developed using the four linear regression methods discussed in Chapter 3 and are compared using five-fold cross-validation. Each regression method was run a first time on either the summer or winter significant variables and then a second time on all variables. The response surface regression method is not performed on all 52 variables due to computational restrictions resulting from the large number of predictor variables. We show that a PLSR model developed using all 52 water and geological variables performed the best with an average 25 percent decrease in RMS error over a simple feed-only model. Detailed results are presented followed by a discussion of various insights that are gained into oil sand screen operation.

4.1 Introduction

It was shown in Chapter 2 that traditional screening variable data is not available in Syncrude's historical databases and collecting a sufficient number of samples to capture plant behaviour is beyond the scope of this work. It was decided, based on suggestions from oil sand screen operators, that water and geological variables should be considered for the development of an advanced statistical regression empirical model. In the next section, we present water and geological time series variables that are recorded by Syncrude. The process of extracting raw data from the databases and cleaning it of erroneous and atypical values is then presented. Once data collection is completed, significant variables are identified using forward stepwise multiple linear regression for both summer and winter screening conditions. An explanatory model is developed showing that the identified significant variables provide additional explained variance of the screening system. Four predictive

models are developed using water and geological variables, and are evaluated using five-fold cross-validation. Each prediction model is then compared against the performance of a feed-only linear regression model. Finally, the significant variables and the explanatory and prediction models are discussed.

4.2 Plant Screen Variables

In this section, we present water and geological variables that have been identified by screen operators as potential reasons for screen performance fluctuation. The identified water and geological variables are measured using online sensors and placed into historical databases. The water variables contain water flow, pressure, and temperature variables. The geological variables provide oil sand physical information such as the ore grade and the percentage of clay.

4.2.1 Water Variables

Water variables describe the screen water system and local weather conditions. In the screen water system, water is added to the oil sand material before reaching the screens and material is sprayed with high pressure water as it passes over the screens. Local weather conditions may also contribute to material moisture content as material can be subjected to rain and high humidity as it moves through the open air sections of the plant. In Section 2.3, high levels of moisture content was shown in previous work [34, 43] to reduce screen performance due to screen plugging and blinding. It is currently unknown how the screening system water usage influences the performance of static screens.

Figure 4.1 shows a diagram of the screen water system and Table 4.1 lists pertinent water system variables measured and placed in a process information database. The location of each sensor in the screen water system in Figure 4.1 is also provided.

In Figure 4.1 we see that water first enters the water system from the extraction facility. At this point sensors collect data on the extraction water flow rate, pressure, and temperature. Further on, a portion of the water is directed to the vortex vessel where it is added to the oil sand material arriving from the surge facility. The flow rate of the water to the vortex is recorded. Over the screens, high pressure water from spray bars is used to help push material through the screen apertures. Unfortunately, at the time of this work, there were no available time series water flow rate or water pressure data for the spray bars. A portion of the water is diverted to the slurry pump glands and time series water pressure data of the diverted water is available. The remaining portion of the water from extraction is directed to the pump box where oil sand is mixed with water to form a slurry that is pumped to extraction. The amount of water that enters the pumpbox is controlled by a valve and time series data is available for the percentage that the valve is open. A portion of the water in the pumpbox is also pumped through a recycle line back to the vortex vessel. The last sensor that is considered in this work records the pressure of the water in the recycle line.

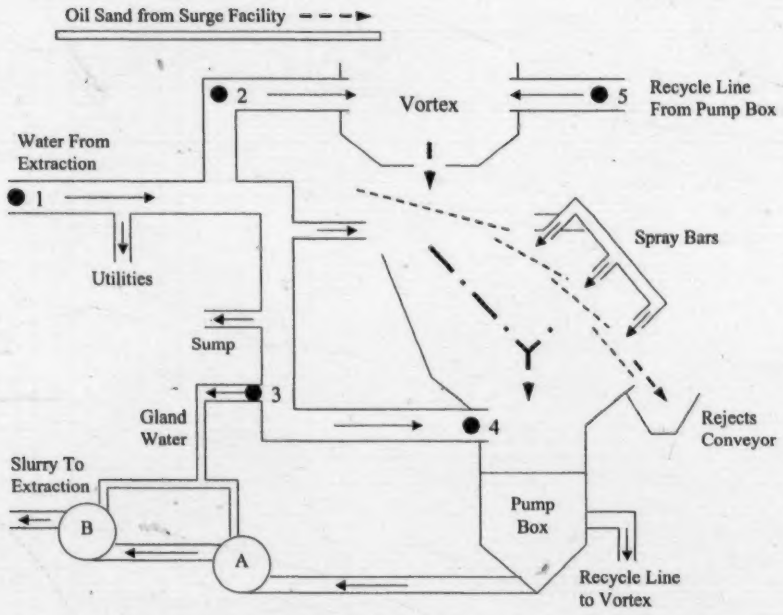

Figure 4.1: Diagram of water flow in the screening system.

Location in Figure 4.1	Variable Name
1	Flow rate of water from extraction
1	Pressure of water from extraction
1	Temperature of water from extraction
2	Flow rate of water to vortex
3	Pressure of water at the pump glands
4	Percent opening of the pumpbox valves
5	Water pressure in the recycle line

Table 4.1: List of water system variables. The first column lists the location in Figure 4.1 where the variable is recorded in the water system.

4.2.2 Geological Variables

Geological properties are known to vary from one area of the mine to the next, and the large variation in oil sand physical properties has been hypothesized by screen operators to affect screen performance. However, it is currently unknown how oil sand geological properties influence screen performance.

Oil sand geological survey data is collected by geologists from core samples taken at 100 metre spacings and entered into a SURPAC geological model [48, 14]. Later, global positioning systems

use the location of the shovel to tag each truck load of oil sand material with the pertinent SURPAC geological information listed in the first three rows of Table 4.2. When the dump trucks deposit their material into one of two toothed roll crushers, NMFB6 or NMFB7, the geological information is saved into the Quality Production Data (QPD) database. Table 4.2 lists the geological variables considered in this work. The first three variables are extracted from the QPD database. The last variable, the material ore grade, is extracted from Syncrude's Process Information (PI) database.

Geological Variables	Unit	Database
Bitumen Content	%	QPD
Clay Content	%	QPD
35 Facies Types	%	QPD
Ore Grade	%	PI

Table 4.2: List of geological variables extracted from the QPD and PI databases.

In Table 4.2, the bitumen content is the approximate percentage by weight of bitumen, a form of crude oil, in the oil sand material at the location where it was dug from the ground. The ore grade variable is the approximate percentage of bitumen in the oil sand exiting the surge facility. The clay content is the approximate percentage of clay in the oil sand material at the location where it was dug from the ground. A facies type describes the lithology and sedimentary textures and structures of oil sand material [44]. In the oil sands industry there are approximately 35 named facies types and there can be any number of different facies types in a given core sample location. Figure 4.2 illustrates the identification of various facies types in a core sample.

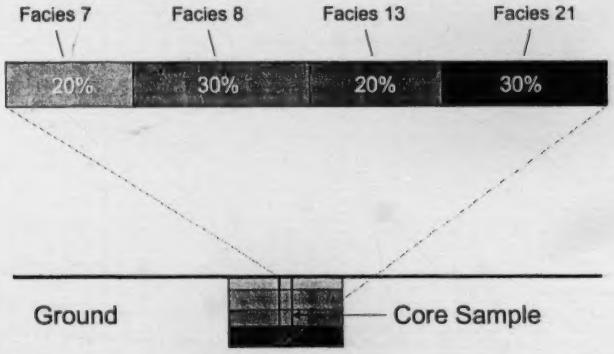

Figure 4.2: The extraction of a core sample and identification of different facies. The percentage value indicates the percentage of the total volume that facies occupies.

At Syncrude, each facies type belongs to one of three categories that are based on the environment in which the facies was formed: marine facies, estuarine facies, or fluvial facies. The facies categories are calculated based on the percentage of each facies in the sample. For example, in Figure 4.2, let us say that Facies 7 and 8 are marine, Facies 13 is estuarine, and Facies 21 is fluvial.

Therefore, in the core sample shown in Figure 4.2, the percentage of marine facies is 50 percent, the percentage of estuarine facies is 20 percent, and the percentage of fluvial facies is 30 percent. The three facies categories are not contained in the QPD database and must be calculated prior to the analysis of the geological data.

4.3 Data Collection

The data collection, integration, and cleaning steps taken to process the raw data are outlined below with careful consideration of the physical aspects of the plant setup. It is known within the oil sands industry that oil sand physical properties vary considerably between summer and winter conditions. To test whether differences exist between seasons, data is collected for summer and winter periods in one minute intervals. It is also noted that operating differences may exist between the two crushers and the two static screens. Consequently, the data is cleaned so that only one crusher and one static screen are in operation for each data point. Raw data is collected for two years (2005 to 2006) of summer screening conditions (May to August) and winter screening conditions (December to March). Due to time restrictions, screening for spring and fall conditions are not considered in this research and are left to future work. Throughout the data cleaning process, assumptions are made regarding system operation so that atypical and erroneous data points are removed.

Data for the 52 water and geological variables, listed in Table 4.3, are gathered from three data sources. The first source is Syncrude's Process Information (PI) database, which contains water sensor data, the material ore grade, and the feed and overflow tonnages. The second source is Syncrude's Quality Production Data (QPD) database, which contains oil sand geological data. The last source is Environment Canada's historical environment website [3], which contains local temperature, humidity, and weather data.

Water and Geological Variables		
Feed Tonnage	Ore Grade	Flow rate of water from extraction
Humidity	35 Facies Types	Pressure of water from extraction
Temperature	Clay Content	Temperature of water from extraction
Weather	Bitumen Content	Flow rate of water to vortex
	Estuarine Content	Pressure of water at the pump glands
	Marine Content	Percent opening of the pump box valves
	Fluvial Content	Water pressure in the recycle line

Table 4.3: List of 52 water and geological variables under study.

In this work, a series of steps are taken to extract, combine, and clean the raw data. The number of data points that were removed at each step set are summarized in Table 4.6 for the Summer data set and in Table 4.7 for the Winter data set.

To start, geological data is extracted from the QPD database. Data points for crushers located in other areas of the plant are deleted. For each data point, the percentage of marine facies, estuarine facies, and fluvial facies are calculated from the 35 different facies types and added to the extracted

data.

The geological data is recorded and time stamped when the material is dumped into the crushers and so must be time shifted forward to correspond with data recorded at the screens. The time shift is calculated as the time for the material to move from the crushers, through the surge pile and to the static screens, and is referred to as the *traversal time*. The material's traversal time is a function of both the amount of material in the surge facility and how quickly the material is removed from the surge pile. The algorithm in Table 4.4 calculates the traversal time for each truck load.

```
1 ·   surgePileCapacity = 10000 tonnes
2 ·   for i = 2 to sizeOfData
3 ·      percentThrough_i = 1 - pileHeightPercentage_i
4 ·      timeElapse = dataTimeStamp_i - dataTimeStamp_{i-1}
5 ·      tonnageExitingSurgePile = feedTonnage_i x timeElapse
6 ·      percentMoveThroughPile = tonnageExitingSurgePile/surgePileCapacity
7 ·      for j = 1 to i-1
8 ·         percentThrough_j = percentThrough_j + percentMoveThroughPile
9 ·         if percentThrough_j >= 100 percent
10 ·           totalTime_j = dataTimeStamp_i - dataTimeStamp_j
11 ·        end if
12 ·     end for
13 ·  end for
```

Table 4.4: Algorithm used to calculate the time taken for a truck load of material to move through the surge pile, or the traversal time. The totalTime$_j$ is the traversal time of truck load j.

The algorithm in Table 4.4 calculates the traversal time based on the surge pile height and the rate at which material exits the surge facility. It is assumed that the capacity of the surge pile is 10000 tonnes. On Line 3, the algorithm calculates the initial position of the truck load in the surge pile based on the surge pile height. The surge pile height is measured using four laser range finders located on the exit end of the crusher conveyor belts, as shown in Figure 4.3. The height is converted to a percentage where 100 percent indicates a full surge pile and zero percent indicates an empty surge pile. The surge pile height for each crusher is assumed to be the mean of the two laser range finders associated with the respective crusher conveyor belt. In Line 4, the elapsed time between the previous truck load $i-1$ and the current truck load i is calculated. Using the elapsed time and the feed tonnage when truck load i is dumped, the approximate amount of material that exited the surge pile is calculated. The feed tonnage when truck load i is dumped is assumed to be the sum of both Train 1 and Train 2 feed tonnages. The percentage the pile moved down, the *incremental percentage through*, is then calculated by dividing the exited material by the total capacity of the surge pile. In lines 7 through 12, the *incremental percentage through* is added to the *percentage through* of each of the previous truck loads that have not finished moving through the surge pile. In lines 9 through 11, the total time is recorded if the total *percentage through* of a truck load reaches 100 percent. The total time for the material to move through the surge pile is the current time minus the time when the truck load entered the surge pile. This process is repeated for each data point in

the QPD data set. Note that the algorithm in Table 4.4 ensures that the order of the data points in the QPD database is kept. This parallels the actual process where the material must exit the surge pile in the order it entered. It is assumed that any traversal time less than 15 minutes or greater than five hours indicates either false data or a temporary plant shutdown. The minimum transversal time of 15 minutes is based on a 2000 tonne surge pile and an 8000 TPH feed tonnage. The maximum traversal time of five hours is based on a 10000 tonne surge pile and a 2000 TPH feed tonnage. Any data point having a traversal time outside of this range is removed.

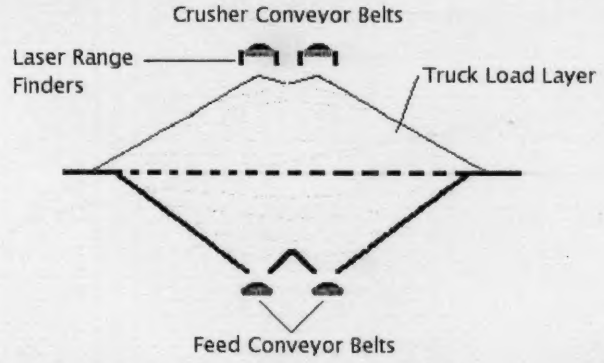

Figure 4.3: Surge pile with two crusher conveyor belts depositing material onto the surge pile and two feed conveyor belts removing material from underneath. Two laser range finders are located on the exit end of each crusher conveyor belt to measure the height of the surge pile.

The data for the Environment Canada *weather* variable, which is entered as text, such as 'rain' or 'clear', are divided into data points with rainy or snowy weather and data points with sunny or clear weather. Data points with rainy or snowy weather are given the value 1, and data points with sunny or clear weather are given the value 0. At this point the QPD data set and the Environment Canada data set are ready to be combined with the PI data set.

The three data sets from the PI database, the QPD database, and Environment Canada's historical environment website are combined for the summer period and then for the winter period. Each of the three data sets has a unique sample frequency. The PI data set is sampled in minute time steps, the QPD data set is sampled in nonuniform time steps that depended on when the truck arrived at the crusher, and the Environment Canada data set is sampled in hour time steps. The QPD dataset and the Environment Canada data set will be up sampled to minute frequency and concatenated with the PI data set. The Environment Canada data set is expanded to minute data through linear interpolation under the assumption that temperature, humidity, and weather move linearly between hourly data points. For the QPD data set, each data point is shifted forward by the previously calculated surge pile traversal time. Every QPD data point is then assigned to the nearest chronological data point in the PI data set. If the nearest chronological data point is greater than 15 minutes, then the data

point in the PI data base is assumed to have no ore data and is removed. Figures 4.4 and 4.5 show the time difference in minutes between each PI data point and the nearest ore data point. For spikes that show a period greater than one day, the period in which the spike occurred is recorded in Table 4.5. At this point 354240 data points are in the Summer data set and 348480 data points are in the Winter data set.

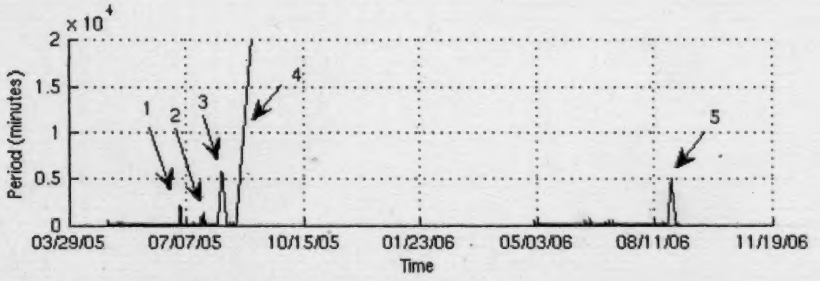

Figure 4.4: Plot of the time difference in minutes between the PI data point and the nearest QPD ore data point for the Summer data set.

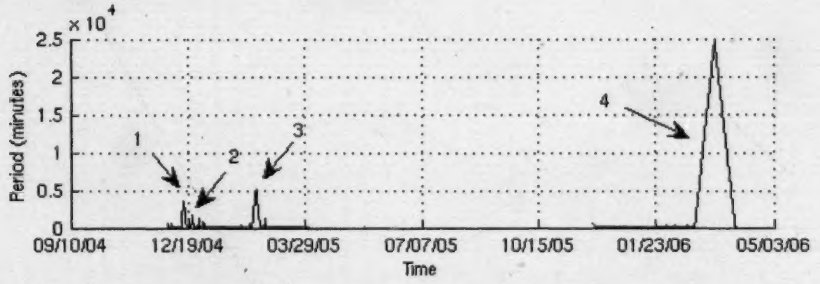

Figure 4.5: Plot of the time difference in minutes between the PI data point and the nearest QPD ore data point for the Winter data set.

Summer	Period	Description
1	Jun 30, 2005 - Jul 03, 2005	NMFB6 and NMFB7 not in operation
2	Jul 21, 2005 - Jul 22, 2005	NMFB6 and NMFB7 not in operation
3	Aug 02, 2005 - Aug 10, 2005	NMFB6 and NMFB7 not in operation
4	Aug 18, 2005 - Aug 31, 2005	NMFB6 and NMFB7 not in operation
5	Aug 22, 2006 - Aug 29, 2006	NMFB6 and NMFB7 not in operation
Winter	**Period**	**Description**
1	Dec 12, 2004 - Dec 17, 2004	NMFB6 and NMFB7 not in operation
2	Dec 21, 2004 - Dec 23, 2004	NMFB6 and NMFB7 not in operation
3	Feb 11, 2005 - Feb 18, 2005	NMFB6 and NMFB7 not in operation
4	Feb 25, 2006 - Mar 31, 2006	NMFB6 and NMFB7 not in operation

Table 4.5: Periods in Figures 4.4 and 4.5 that are longer than one day.

The next step removes all data points for when both crushers are operating simultaneously. The

QPD data set is traversed with a 30 minute window centred around the data point of interest. If a crusher different from that of the current data point is in operation during the 30 minute window, the current data point is marked for removal. The 30 minute window reflects a 15 minute uncertainty in the material surge pile traversal time and is chosen based on a 10 percent approximate error in the surge pile height measurements. A 10 percent error in surge pile height for a 10000 tonne surge pile and a low feed tonnage rate of 2500 TPH results in 24 minutes of possible error in traversal time. An additional six minutes of uncertainty is included to account for uncertainty in the feed tonnage weightometer reading and the time to travel from the crusher to the surge facility and the surge facility to the screen. All data points having two crushers in operation within a 30 minute window are removed from the PI data set. Figures 4.6 and 4.7 show histograms of the period length between a data point recorded at one crusher and a data point recorded at the second crusher for the the Summer and Winter data sets respectively. Period lengths greater than 15 minutes are placed in the 15 minute bin. Approximately 180 thousand data points from the summer period and 160 thousand data points from the winter period are recorded when only one crusher is in operation.

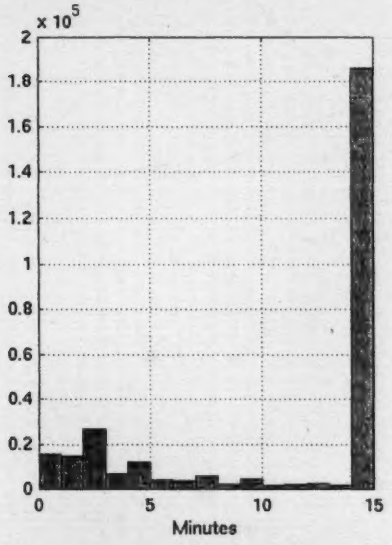

Figure 4.6: Histogram of the period lengths between a data point recorded at one crusher and a data point recorded at the second crusher for the Summer data set. Period lengths greater than 15 minutes are placed in the 15 minute bin.

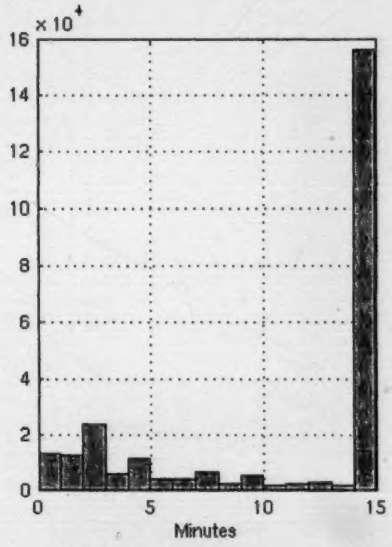

Figure 4.7: Histogram of the period lengths between a data point recorded at one crusher and a data point recorded at the second crusher for the Winter data set. Period lengths greater than 15 minutes are placed in the 15 minute bin.

When the reject reprocessing system is not in operation, screen overflow material bypasses the overflow conveyor belt and is dumped on the ground. Consequently, the weightometer measuring the tonnage of overflow material, located on the overflow conveyor belt that carries material from the screens to the rejects reprocessing system, is recording values for an empty conveyor belt. To ensure

that the rejects reprocessing system is in operation and that the overflow tonnage measurement is valid, a status flag in the PI database is checked for each data point. A flag value of one indicates the rejects reprocessing system was running and a value of zero indicates it was not running. All data points when the status flag indicates rejects reprocessing is not running are removed. Figures 4.8 and 4.9 show histograms of the number of data points for the summer and winter periods for each value of the status flag. Approximately 130 thousand data points from the summer period and 100 thousand data points from the winter period have valid overflow tonnage measurements.

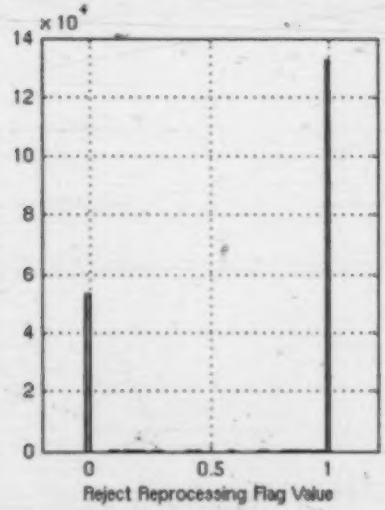

Figure 4.8: Histogram of the status flag values for the Summer data set.

Figure 4.9: Histogram of the status flag values for the Winter data set.

To ensure that the overflow tonnage is within typical operating ranges, histograms of the overflow tonnage data are plotted in Figures 4.10 and 4.11 for summer and winter periods. It was decided that data points in the Summer and Winter data sets having an overflow tonnage less than 100 TPH are assumed to be invalid and are removed. Figure 4.10 shows that approximately 1500 data points have an overflow tonnage value less than 100 TPH in the Summer data set and Figure 4.11 shows that approximately 100 data points have an overflow tonnage less than 100 TPH in the Winter data set.

To ensure that the feed tonnage is within typical operating ranges, histograms of the feed tonnage data is plotted in Figures 4.12 and 4.13 for summer and winter periods. Note that in Figures 4.12 and 4.13 bins for feed tonnages less than 250 TPH are not included as they are considered to represent the system when it is not in operation. It was decided that data points in the Summer and Winter data sets having a feed tonnage between 2000 TPH and 8000 TPH are considered to be recorded during typical operation.

The summer and Winter data sets are now subdivided based on which train is in operation. The

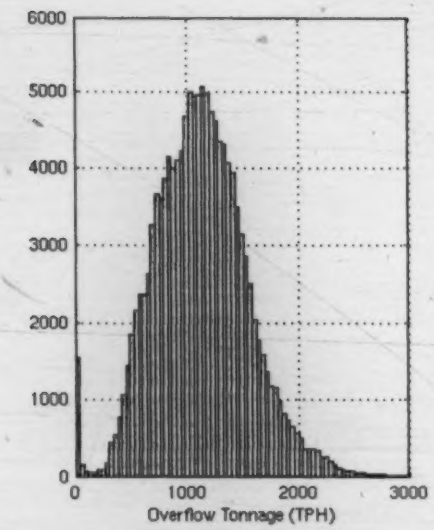

Figure 4.10: Histogram of the overflow tonnage values for the Summer data set.

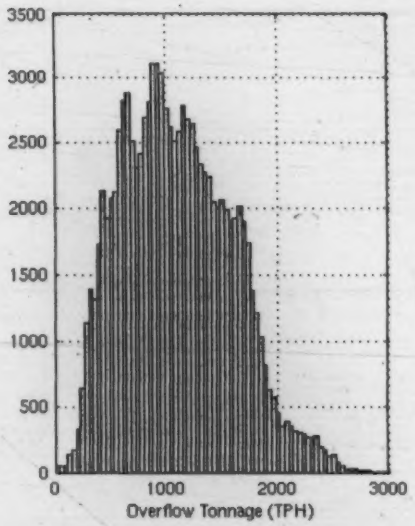

Figure 4.11: Histogram of the overflow tonnage values for the Winter data set.

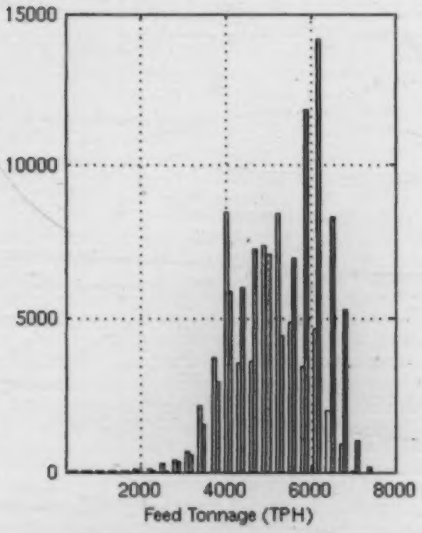

Figure 4.12: Histogram of the feed tonnage values for the Summer data set.

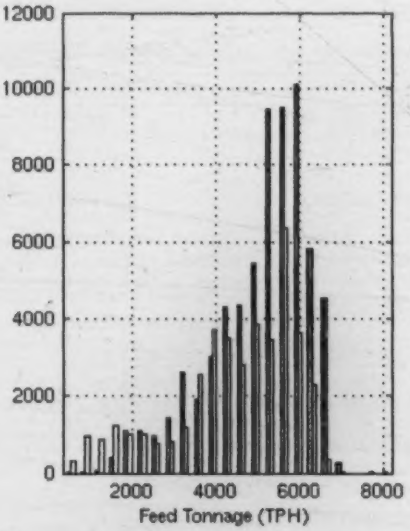

Figure 4.13: Histogram of the feed tonnage values for the Winter data set.

Summer Train 1 data set is compiled of data points where the Train 1 feed tonnage is between 2000 and 8000 TPH and the Train 2 feed tonnage is less than 1 TPH. Similarly, the Summer Train 2 data

set is compiled of data points where the Train 2 feed tonnage is between 2000 and 8000 TPH and the Train 1 feed tonnage is less than 1 TPH. This procedure is repeated for the Winter data set.

Earlier it was remarked that it is observed within the oil sands industry that oil sand properties vary between summer and winter months. For this reason, May through August was selected to represent the summer period and December through March was selected to represent the winter period. However, a strict adherence of the temperature to these assumed states during the selected months is unlikely. For this reason, each data set is cleaned based on the ambient temperature. Typical temperatures for summer months are assumed to be greater than $5°C$ and, therefore, any data point having an ambient temperature below $5°C$ is removed. Typical temperatures for winter months are assumed to be below -$5°C$ and, therefore, any data point having an ambient temperature above -$5°C$ is removed. Figures 4.14 and 4.15 show the histogram of the ambient temperature for the summer and winter periods.

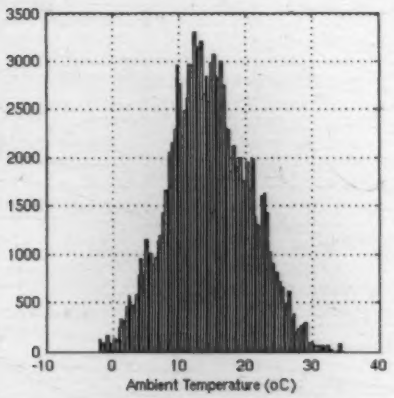

Figure 4.14: Histogram of the temperature values for the cleaned Summer data set.

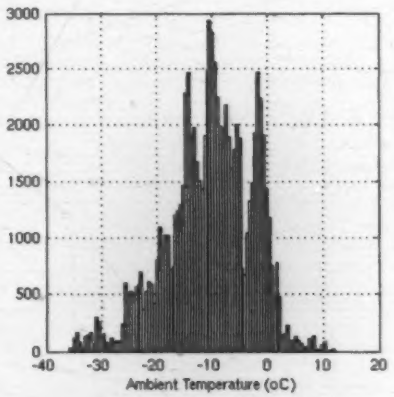

Figure 4.15: Histogram of the temperature values for the cleaned Winter data set.

As data is entered into the PI database from sensors, circumstances, like a plant shutdown, may cause data to fall outside of typical operating range. In these situations the database replaces the invalid data values with tags, such as 'Bad Data' or 'Scan Off'. Data points that were found to contain non-numerical values were removed.

At this point the cleaning of the data is complete. Tables 4.6 and 4.7 list the number of data points removed at each stage for the summer months and the winter months respectively. Note that the cleaning steps were applied serially and the number of data points that were removed in each step rely on the order the steps were performed. A reordering of the steps would result in a change in the number of data points removed at each step.

Tables 4.6 and 4.7 show that after cleaning 32.7 percent of the Summer data set and 18.2 percent of the Winter data set is usable for analysis of the screening system.

Summer Data Set				
Initial number of data points	354240		100.0%	
PI Data points that don't have ore data	64675		18.3%	
C1 and C2 in operation at same time	102424		28.9%	
Rejects reprocessing not operating	54312		15.3%	
Overflow tonnage less than 100 TPH	1741		0.5%	
Feed tonnage outside 2000 and 8000 TPH	7828		2.2%	
Divide data: Train 1 and Train 2	**Train 1**		**Train 2**	
Initial number of data points for each train	46774	13.2%	76486	21.6%
Temperature values less than 5°C	5645	1.6%	1606	0.5%
Erroneous data	0	0.0%	7	0.0%
Number of data points after cleaning	41129	11.6%	74873	21.1%

Table 4.6: The number of data points removed at each stage of the data cleaning process for the Summer data set. C1 and C2 refer to crushers NMFB6 and NMFB7 respectively.

Winter Data Set				
Initial number of data points	348480		100.0%	
PI Data points that don't have ore data	96843		27.8%	
C1 and C2 in operation at same time	93063		26.7%	
Rejects reprocessing not operating	57629		16.5%	
Overflow tonnage less than 100 TPH	133		0.0%	
Feed tonnage outside 2000 and 8000 TPH	16408		4.7%	
Divide data: Train 1 and Train 2	**Train 1**		**Train 2**	
Initial number of data points in each train	53983	15.5%	30421	8.7%
Temperature values greater than -5°C	15077	4.3%	5664	1.6%
Erroneous data	76	0.0%	60	0.0%
Number of data points after cleaning	38830	11.1%	24697	7.1%

Table 4.7: The number of data points removed at each stage of the data cleaning process for the Winter data set. C1 and C2 refer to crushers NMFB6 and NMFB7 respectively.

The data is further divided into subsets when only one crusher and one screen are in operation. The resulting number of data points for each permutation of crusher and screen are listed in Table 4.8. Note that the terms Crusher 1 and Crusher 2 refer to crushers NMFB6 and NMFB7 respectively. In Table 4.8, the data is further grouped into 30 minute periods. The periods reflect the uncertainty in the accuracy of the cleaned data. A median filter is applied to each group to obtain one representative data point. When grouping the data points, a minimum data point density of 75 percent, or 23 data points, in each group is required.

		Summer Data Set		Winter Data Set	
		Train 1	**Train 2**	**Train 1**	**Train 2**
Original	Crusher 1	37624	5725	20753	733
	Crusher 2	3505	69148	18077	23964
Grouped	Crusher 1	1222	166	629	15
	Crusher 2	101	2239	573	767

Table 4.8: The number of data points for each data set without grouping and the number of data points in each data set when grouped using a 30 minute median filter. The 30 minute median filter required a minimum of 23 data points for each group. Crusher 1 and Crusher 2 refer to crushers NMFB6 and NMFB7 respectively.

Table 4.8 shows that the number of data points in the summer months for when Train 1 & Crusher 1 and Train 2 & Crusher 2 are in operation are significantly higher than when Train 1 & Crusher 2 and Train 2 & Crusher 1 are in operation. In the winter months, only the number of data points for when Train 2 & Crusher 1 are in operation are significantly lower than the remaining three configurations. To simplify the analysis, only data points for when Train 1 & Crusher 1 and Train 2 & Crusher 2 are in operation will be used. The combined Train 1 & Crusher 1 system will be referred to as Train 1 and the combined Train 2 & Crusher 2 system will be referred to as Train 2.

The data set naming convention used for the analysis of the data is given in Table 4.9.

Data Set Name	Data Set Description
Summer Data Set	All data during the summer months
Winter Data Set	All data during the winter months
Summer Train 1 Data Set	All data when Train 1 is in operation during the summer months
Summer Train 2 Data Set	All data when Train 2 is in operation during the summer months
Winter Train 1 Data Set	All data when Train 1 is in operation during the winter months
Winter Train 2 Data Set	All data when Train 2 is in operation during the winter months

Table 4.9: Data set naming convention used in the data analysis portion of this work.

In this section, through a series of data collection, integration, and cleaning steps based on an understanding of plant operation, water and geological data were prepared for analysis.

4.4 Outline of Analysis

A series of analyses are carried out in this work to identify significant variables and develop explanatory and prediction oil sand screen models. The goal of this work is to gain as much information as possible from the cleaned data in order to provide the screen operators with insight into how the screens behave and a preliminary tool with which to predict what may happen when the screening environment changes.

An interesting question often posed regarding oil sand screening is which variables of those considered are the most significant. If these variables were known, then steps could be taken to adjust screen operation accordingly. Forward stepwise multiple linear regression (FSMLR) is used to identify a preliminary set of significant screen variables for both summer and winter screening conditions. FSMLR initially starts with zero variables in the model and iteratively adds to the model the variable, taken from the set of unused variables, that provides the greatest explanation of the remaining system variance. FSMLR is stopped when the reduction in RMS error of an additional variable is less than 3 TPH. An explanatory MLR model of the screening system is then created using the significant variables and studied using coefficient analysis to gain insight into the screening process. The resulting explanatory model is evaluated using p-values, standard error, and variance inflation factors to ensure that model is statistically significant.

Once an explanatory model is completed, more powerful regression methods can be used to cre-

ate models with greater predictive accuracy. Four common regression methods were presented in Chapter 3: multiple linear regression (MLR), response surface regression (RSR), principle component regression (PCR), and partial least squares regression (PLSR). Each regression method is run first on either the summer or winter significant variables and then a second time on all variables. The response surface regression method is not run on all 52 variables due to computational restrictions resulting from the large number of predictor variables. To put into context the performance of the prediction models, a simple linear regression (SLR) model is also developed. SLR will use only the feed tonnage as input variable and will provide a baseline in performance that the other regression models can be compared against. The resulting models are evaluated using five-fold cross-validation and the average RMS error and adj-R^2 values are compared. A discussion of the resulting prediction model performances is presented in Section 4.6.

4.5 Results

In this section the results of the analytical procedure are presented. Significant variables are identified for both summer and winter screening conditions using FSMLR. An explanatory model is developed based on the identified significant variables and validated using measures of multicollinearity and statistical significance. Finally, seven prediction models are developed using MLR, RSR, PCR, and PLSR and compared to an SLR model that uses only the feed tonnage as input variable. We show that a partial least squares regression prediction model using all 52 water and geological variables performed the best with an average 25 percent decrease in RMS error over a simple feed-only model.

4.5.1 Explanatory Model Development

FSMLR is used to identify significant variables for both summer and winter screening conditions using the cleaned data. The resulting significant variables are listed in Table 4.10, along with the resulting RMS error and adj-R^2 values, in the order in which they were identified.

Summer Data Set			Winter Data Set		
Variable Name	**RMS Error**	**Adj-R^2**	**Variable Name**	**RMS Error**	**Adj-R^2**
Feed Tonnage	315 TPH	0.21	Pumpbox % Open	335 TPH	0.28
Pumpbox % Open	293 TPH	0.32	Water Temperature	310 TPH	0.38
Ore Grade	277 TPH	0.39	Feed Tonnage	287 TPH	0.47
Facies 21	271 TPH	0.41	Gland Water kPa	274 TPH	0.51
Facies 8	268 TPH	0.43	Recycle Line kPa	265 TPH	0.55
			Vortex Flow	261 TPH	0.56

Table 4.10: List of the significant variables for summer and winter screening conditions.

From Table 4.10, we see that five variables were found to be significant for summer screening conditions and six variables were found to be significant for winter screening conditions. Notice that the significant summer variables are composed primarily of geological variables and the winter

significant variables are composed primarily of water variables. This result supports the observation that differences exist between summer and winter screening behaviour and that two separate screening models are needed for each period.

For each of the Train 1 and Train 2 screening conditions for summer and winter, four explanatory models are created and evaluated to understand how significant the identified variables are. The first model, SLR, is a simple linear regression model using only the feed tonnage as input. The second model, MLR_{sig}, is an MLR model that uses the significant variables in Table 4.10 for summer or winter screening conditions. The third model, MLR_{all}, uses all 52 water and geological variables. The last model, MLR_{comb}, is a combined model for both Train 1 and Train 2, where the same coefficients are used. This last model is used to determine whether a combined model for both Train 1 and Train 2 is possible. Looking first at the summer data set, Table 4.11 lists each of the four summer models for both the Train 1 and Train 2 data sets.

Model	Summer Train 1		Summer Train 2	
	RMSE	Adj-R^2	RMSE	Adj-R^2
SLR	262 TPH	0.51	312 TPH	0.16
MLR_{sig}	244 TPH	0.58	274 TPH	0.35
MLR_{all}	223 TPH	0.65	252 TPH	0.45
MLR_{comb}	253 TPH	0.55	276 TPH	0.35

Table 4.11: Models used to determine how significant the identified summer screen variables are for summer screening conditions. The first model, SLR, is a simple linear regression model using only the feed tonnage as input. The second model, MLR_{sig}, uses only the summer significant variables in Table 4.10. The third model, MLR_{all}, uses all water and geological variables. The last model, MLR_{comb}, a model that uses a common set of variable coefficients for both the Train 1 and Train 2 data sets. It is used to determine whether a combined model for both Train 1 and Train 2 is possible.

Table 4.11 shows that MLR_{sig} performs, on average, 10 percent better than SLR, as measured by the RMS error. Furthermore, MLR_{all} performs, on average, 17 percent better than SLR, as measured by the RMS error. This shows that using the five summer significant variables provides roughly 60 percent of the total reduction in RMS error achieved by using all 52 water and geological variables. This result indicates that the selected summer variables are significant in reducing the residual error and increasing the amount of explained system variance for summer screening conditions. When using a common set of coefficients for the significant summer variables for both Train 1 and Train 2 it was found that the RMS error increased by 2 percent. This result indicates that a common model is applicable for both Train 1 and Train 2 during summer screening conditions.

Table 4.12 shows that MLR_{sig} performs, on average, 32 percent better than SLR, as measured by the RMS error. Furthermore MLR_{all} performs, on average, 38 percent better than SLR, as measured by the RMS error. This shows that using the six winter significant variables provides roughly 84 percent of the total reduction in RMS error achieved by using all 52 water and geological variables. This result indicates that the selected winter variables are significant in reducing the residual error and increasing the amount of explained system variance for winter screening conditions. When

Model	Winter Train 1		Winter Train 2	
	RMSE	Adj-R^2	RMSE	Adj-R^2
SLR	409 TPH	0.24	273 TPH	0.05
MLR_{sig}	298 TPH	0.60	173 TPH	0.62
MLR_{all}	279 TPH	0.65	150 TPH	0.71
MLR_{comb}	324 TPH	0.57	199 TPH	0.55

Table 4.12: Models used to determine how significant the identified winter screen variables are for winter screening conditions. The first model, SLR, is a simple linear regression model using only the feed tonnage as input. The second model, MLR_{sig}, uses only the winter significant variables in Table 4.10. The third model, MLR_{all}, uses all water and geological variables. The last model, MLR_{comb}, a model that uses a common set of variable coefficients for both the Train 1 and Train 2 data sets. It is used to determine whether a combined model for both Train 1 and Train 2 is possible.

using a common set of coefficients for the significant winter variables for both Train 1 and Train 2 it was found that the RMS error increased by 12 percent. This result indicates that a common model is not applicable for both Train 1 and Train 2 during winter screening conditions. Figures 4.16 and 4.17 plot the winter data for Train 1 and Train 2 respectively.

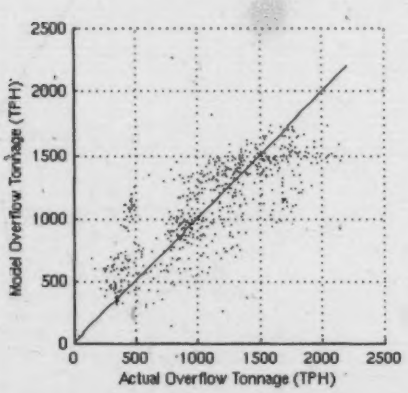

Figure 4.16: Model overflow tonnage versus the actual overflow tonnage for the Winter Train 1 data set.

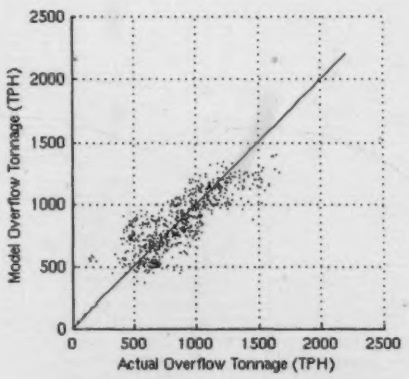

Figure 4.17: Model overflow tonnage versus the actual overflow tonnage for the Winter Train 2 data set.

From Figures 4.16 and 4.17, we see that there is a noticeable difference between the Train 1 screening system and the Train 2 screening system that is not captured in the list of variables considered in this work. Further investigation into reasons for the performance difference between the Train 1 and Train 2 screening systems is beyond the scope of this research and is left for future work.

The list of significant variables in Table 4.10 contains three variables associated with ore characteristics for the summer period and five variables associated with water variables for the winter period. The next two sections present data from the Summer Train 2 data set and the Winter Train 2 data set that underline the significance of considering geological variables for modelling summer screening performance and water variables for modelling winter screening performance.

Significance of Ore Variables For Summer Screening

To underline the importance of using geological variables during the summer period, two linear regression models are compared that use two different sets of variables. The models are developed using the Summer Train 2 data set. The first model uses only the feed tonnage, which is typically used as the primary means of modelling the overflow tonnage. The second model uses the significant summer variables listed in Table 4.10. The pumpbox valve percent open variable is not included in order to illustrate the importance of the geological variables. The results of the two models are presented in Table 4.13.

Regression Model	RMS Error	Adj-R^2
SLR: Feed Tonnage	312 TPH	0.16
MLR: Significant Geological Variables	275 TPH	0.35

Table 4.13: The results of two regression models that use two different sets of variables on the Summer Train 2 data set. The average overflow tonnage during the period is 1076 TPH. The SLR model uses only the feed tonnage as input variable. The MLR model uses the identified significant geological variables, including the feed tonnage variable, listed in Table 4.10. The pumpbox valve percent open variable is not included in order to underline the importance of the geological variables.

Table 4.13 shows that including the significant geological variables in the MLR model results in a 19 percent decrease in RMS error over the SLR model. Figures 4.18 and 4.19 show the model overflow tonnage versus the actual overflow tonnage for the SLR model and the MLR model respectively.

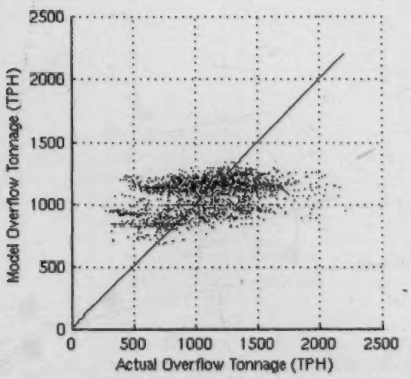

Figure 4.18: Model overflow tonnage versus the actual overflow tonnage for the SLR model for the Summer Train 2 data set.

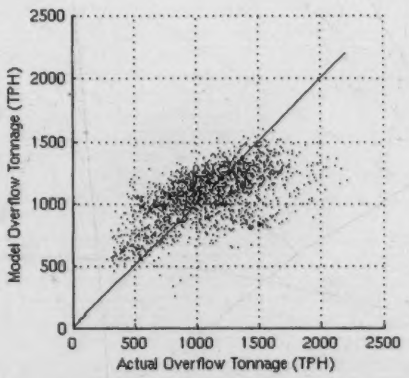

Figure 4.19: Model overflow tonnage versus the actual overflow tonnage for the MLR model for the Summer Train 2 data set.

The results presented in Table 4.13 underline the significance of using geological variables to model summer screening performance.

Significance of System Water Variables For Winter Screening

To underline the importance of using water variables to model screen performance during the winter period, two linear regression models are compared that use two different sets of variables. The models are developed using the Winter Train 2 data set. The first model uses only the feed tonnage, which is typically used as the primary means of modelling the overflow tonnage. The second model uses the significant water variables listed in Table 4.14. The results of the two models are presented in Table 4.14.

Regression Model	RMS Error	Adj-R^2
SLR: Feed Tonnage	273 TPH	0.05
MLR: Significant Water Variables	171 TPH	0.63

Table 4.14: The results of two regression models that use two different sets of variables on the Winter Train 2 data set. The average overflow tonnage during the period is 855 TPH. The SLR model uses only the feed tonnage as input variable. The MLR model uses the identified significant water variables listed in Table 4.10.

Table 4.14 shows that including the significant water variables in the MLR model results in a 39 percent decrease in RMS error over the SLR model. Figures 4.20 and 4.21 show the model overflow tonnage versus the actual overflow tonnage for the SLR model and the MLR model respectively.

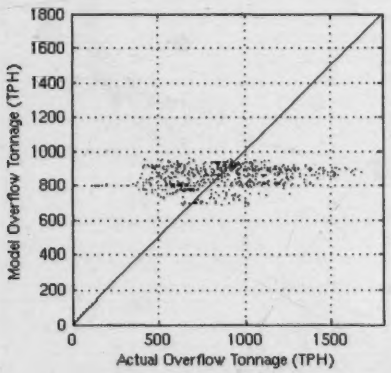

Figure 4.20: Model overflow tonnage versus the actual overflow tonnage for the SLR model for the Winter Train 2 data set.

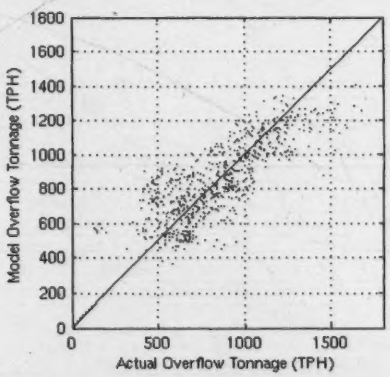

Figure 4.21: Model overflow tonnage versus the actual overflow tonnage for the MLR model for the Winter Train 2 data set.

The results presented in Table 4.14 underline the significance of using water variables to model winter screening overflow tonnage.

Explanatory Model Development

The development of an explanatory model provides insight into the screening system behaviour. The explanatory model is composed of one submodel for summer screening conditions, and two submodels for winter Train 1 screening conditions and winter Train 2 screening conditions. The

following section presents multiple linear regression explanatory submodels for each screening condition along with evaluation measures that ensure the results are statistically significant.

Summer Screen Submodel					
Variable	**Coef.**	**Stand. Err.**	**VIF**	**T for H0**	**Prob > |T|**
Feed Tonnage	0.211	0.0055	1.21	38	<0.0001
Ore Grade	58.73	2.9	1.22	21	<0.0001
Facies 8	8.72	0.42	1.05	21	<0.0001
Facies 21	-2.62	0.27	1.03	-9.8	<0.0001
Pumpbox Valve % Open	19.42	1.7	1.06	11	<0.0001
Constant	-1248	52	0.00	-24	<0.0001

Table 4.15: Coefficient values and statistical evaluation measures for the summer screen submodel.

In Table 4.15 we see that for all variables in the summer screen submodel the VIF is close to one and the p-value is very small indicating the each variable is significant in the presence of the remaining variables. The calculated f-value of the summer screen submodel is 520 indicating a statistically significant submodel.

Winter Train 1 Screen Submodel					
Variable	**Coef.**	**Stand. Err.**	**VIF**	**T for H0**	**Prob > |T|**
Feed Tonnage	0.187	0.013	1.85	14	<0.0001
Extraction Water Temperature	23.62	4.5	1.42	4.9	<0.0001
Flow rate of water to vortex	-1.42	0.20	2.41	-7.0	<0.0001
Pump Gland Water Pressure	-2.42	0.31	2.69	-7.7	<0.0001
Pumpbox Valve Percentage Open	12.65	2.0	1.60	6.2	<0.0001
Recycle Line Water Pressure	-13.35	1.2	1.67	-11	<0.0001
Constant	959	550	0.00	1.8	0.0805

Table 4.16: Coefficient values and statistical evaluation measures for the winter Train 1 screen submodel.

In Table 4.16 we see that for all variables in the winter Train 1 screen submodel the VIF is close to one. Furthermore, we see that the significant variables are statistically significant, except for the constant, which is only significant at the 8 percent level. At the 5 percent level, the p-value for the constant shows it can be ignored without series loss in performance. The calculated f-value of the winter Train 1 screen submodel is 154 indicating a statistically significant submodel.

Winter Train 2 Screen Submodel					
Variable	**Coef.**	**Stand. Err.**	**VIF**	**T for H0**	**Prob > |T|**
Feed Tonnage	0.0765	0.0078	1.71	9.8	<0.0001
Extraction Water Temperature	33.41	1.8	1.29	18	<0.0001
Flow rate of water to vortex	-0.89	0.092	2.08	-9.7	<0.0001
Pump Gland Water Pressure	-0.17	0.15	3.01	-1.2	0.2428
Pumpbox Valve Percentage Open	14.51	1.6	3.27	9.0	<0.0001
Recycle Line Water Pressure	-7.03	0.75	3.42	-9.4	<0.0001
Constant	-1449	210	0.00	-6.9	<0.0001

Table 4.17: Coefficient values and statistical evaluation measures for the winter Train 2 screen submodel.

In Table 4.17 we see that for all variables in the winter Train 2 screen submodel the VIF is close to one. Furthermore, we see that all variables except the pump gland water pressure are statistically significant. The pump gland water pressure was found to be significant in the Winter Train 1 data set but not in the Winter Train 2 data set. For this reason the pump gland water pressure can be ignored for the Train 2 screen submodel. The calculated f-value of the winter Train 2 screen submodel is 215 indicating a statistically significant submodel.

In this section, significant variables were identified for summer and winter periods. The summer significant variables were found to be composed primarily of geological variables, whereas the winter significant variables were found to be composed primarily of water variables. Furthermore, the Summer Train 1 and Train 2 data sets were found to be well modelled by the same set of coefficients, however, the Winter Train 1 and Train 2 data sets were not. A final explanatory model is therefore composed of three submodels, a submodel for the summer period and two submodels for the winter Train 1 period and the winter Train 2 period. In the next section the three submodels are extended by more advanced regression methods.

4.5.2 Prediction Model Development

In the previous section we looked at developing an explanatory model, composed of three submodels, using identified significant variables. The three submodels included a summer screening submodel, a winter Train 1 submodel, and a winter Train 2 submodel. More advanced regression methods were not considered for explanatory model development as they are ill-suited for gaining insight into the underlying system. However, in this section, the goal is to develop a model that minimizes the RMS error and maximizes the explained variance. For this reason, more advanced regression methods can be investigated.

Four advanced linear regression methods, overviewed in Chapter 3 and listed in Table 4.18, are evaluated for the development of a preliminary prediction model. Each regression method is run a first time on the summer or winter significant variables, and then a second time with all 52 water and geological variables. The response surface regression method is not run on all 52 variables due to computational restrictions resulting from the large number of predictor variables.

Five-fold cross-validation is used to evaluate the resulting models. For five-fold cross-validation, the Summer Train 1, Summer Train 2, Winter Train 1, and Winter Train 2 data sets are each divided into five groups using random selection. Four of the groups are used to train a model, and the fifth is used to evaluate the new model. The previous step is repeated, setting a different group as the test group for each iteration, until each group has been used as a test group. The mean and standard deviation of the five results are recorded. The Summer Train 1 and Summer Train 2 data sets are used in place of the Summer data set to further evaluate the regression methods in Table 4.18. The data sets are divided into five groups in place of 10 or 20 groups to ensure that the test group was sufficiently large for proper evaluation of the model's performance. For PCR and PLSR,

Method	Name	Description
SLR	Simple Linear Regression	Using Only Feed rate
MLR	Multiple Linear Regression	Using Significant Variables
RSR	Response Surface Regression	Using Significant Variables
PCR	Principle Component Regression	Using Significant Variables
PLSR	Partial Least Squares Regression	Using Significant Variables
MLRa	Multiple Linear Regression	Using All Variables
PCRa	Principle Component Regression	Using All Variables
PLSRa	Partial Least Squares Regression	Using All Variables

Table 4.18: A list of regression methods used to develop prediction models. Each regression method is run twice, a first time on the summer or winter significant variables, and then a second time with all 52 water and geological variables. The response surface regression method is not run on all 52 variables due to computational restrictions resulting from the large number of predictor variables.

the percentage of explained variance of the data set that was represented by components or latent vectors was thresholded at 85 percent.

In the following four tables, showing the results from each prediction model, the first column gives the name of the model. The second and third columns give the mean and standard deviation of the RMS error and adj-R^2 of each model. Since the results of each model will shift up or down together based on the randomly selected groups of data, the last two columns show the relative performance, as measured by the difference in RMS error and the adj-R^2, between the best performing model and each of the remaining models. Table 4.19 shows the method used to calculate the relative performance between the best performing model, b, and a second model, i. The method in Table 4.19 is repeated for each of the remaining models. Table 4.20, lists the results for the Summer Train 1 data set.

1 · for j = 1 to M = 5
2 · T = [T , RMS_b - RMS_i]
3 · end for
4 · RMS_{mean} = mean(T)
5 · RMS_{std} = std(T)

Table 4.19: Method used to calculate the relative performance between the best performing model and the remaining models. M is the total number of randomly selected groups of data in the cross-validation and b is the best performing model.

Table 4.20 shows that when only the significant variables were considered, RSR has the best performance with a 10 percent lower RMS error than the SLR model. When all variables were included, the PLSRa model had the best performance of both groups with a 12 percent decrease in RMS error over the SLR model. However the PLSRa model did not perform significantly better than the RSR model. However, PLSRa did perform on average better than RSR and, for this reason, is selected as the top performing prediction model. In all cases, the regression methods that used the water and geological variables performed significantly better than the feed-only SLR model. The process is repeated using the Summer Train 2 data set and Table 4.21 lists the results.

Model	Model Performance		PLSRa	
	RMSE	Adj-R^2	RMSE	Adj-R^2
SLR	263 ± 20 TPH	0.51 ± 0.04	30±8 TPH	0.11 ± 0.03
MLR	247 ± 15 TPH	0.57 ± 0.05	15±6 TPH	0.05 ± 0.02
RSR	237 ± 13 TPH	0.60 ± 0.05	4±9 TPH	0.01 ± 0.03
PCR	249 ± 13 TPH	0.56 ± 0.05	17±7 TPH	0.06 ± 0.02
PLSR	247 ± 13 TPH	0.57 ± 0.05	14±7 TPH	0.05 ± 0.02
MLRa	244 ± 13 TPH	0.58 ± 0.04	12±3 TPH	0.04 ± 0.01
PCRa	263 ± 14 TPH	0.51 ± 0.05	31±9 TPH	0.12 ± 0.03
PLSRa	232 ± 12 TPH	0.62 ± 0.04	—	—

Table 4.20: Results for each model for the Summer Train 1 data set. The relative performance between the best performing model and the remaining models is provided in the last two columns. The best performing model for the Summer Train 1 data set is the PLSRa model. The overflow tonnage for the Summer Train 1 data set is 1142±375 TPH.

Model	Model Performance		PLSRa	
	RMSE	Adj-R^2	RMSE	Adj-R^2
SLR	312 ± 9 TPH	0.16 ± 0.03	56±7 TPH	0.27 ± 0.03
MLR	276 ± 12 TPH	0.34 ± 0.04	19±5 TPH	0.09 ± 0.02
RSR	265 ± 12 TPH	0.39 ± 0.04	8±7 TPH	0.03 ± 0.03
PCR	276 ± 12 TPH	0.34 ± 0.04	32±6 TPH	0.15 ± 0.03
PLSR	276 ± 11 TPH	0.34 ± 0.03	19±5 TPH	0.09 ± 0.02
MLRa	265 ± 9 TPH	0.39 ± 0.04	8±2 TPH	0.04 ± 0.01
PCRa	291 ± 10 TPH	0.27 ± 0.04	34±6 TPH	0.16 ± 0.02
PLSRa	257 ± 8 TPH	0.43 ± 0.03	—	—

Table 4.21: Results for each model for the Summer Train 2 data set. The relative performance between the best performing model and the remaining models is provided in the last two columns. The best performing model for the Summer Train 2 data set is the PLSRa model. The overflow tonnage for the Summer Train 2 data set is 1076±340 TPH.

Table 4.21 shows that when only the significant variables were considered, RSR had the best performance with a 15 percent lower RMS error than the SLR model. When all variables were included, the PLSRa model had the best performance of both groups with an 18 percent decrease in RMS error over the SLR model. In all cases, the regression methods that used the water and geological variables performed significantly better than the feed-only SLR model. The process is repeated using the Winter Train 1 data set and Table 4.22 lists the results.

Table 4.22 shows that when only the significant variables were considered, PLSR had the best performance with a 26 percent lower RMS error than the SLR model. When all variables were included, the PLSRa model had the best performance of both groups with a 27 percent decrease in RMS error over the SLR model. Note that PLSRa did not perform statistically better than the MLR model or the PLSR model. However, PLSRa did perform on average better than MLR or PLSR and, for this reason, is selected as the top performing prediction model. In all cases, the regression methods that used the water and geological variables performed significantly better than the feed-only SLR model. The process is repeated using the Winter Train 2 data set and Table 4.23 lists the results.

Model	Model Performance		PLSRa	
	RMSE	Adj-R^2	RMSE	Adj-R^2
SLR	412 ± 28 TPH	0.24 ± 0.09	110±20 TPH	0.35 ± 0.06
MLR	306 ± 22 TPH	0.57 ± 0.06	9±11 TPH	0.03 ± 0.03
RSR	325 ± 35 TPH	0.55 ± 0.11	26±23 TPH	0.07 ± 0.06
PCR	310 ± 21 TPH	0.57 ± 0.07	21±14 TPH	0.06 ± 0.04
PLSR	305 ± 21 TPH	0.58 ± 0.07	6±11 TPH	0.02 ± 0.03
MLRa	330 ± 20 TPH	0.51 ± 0.06	30±5 TPH	0.08 ± 0.02
PCRa	349 ± 21 TPH	0.45 ± 0.07	49±16 TPH	0.15 ± 0.05
PLSRa	300 ± 17 TPH	0.59 ± 0.05	—	—

Table 4.22: Results for each model for the Winter Train 1 data set. The relative performance between the best performing model and the remaining models is provided in the last two columns. The best performing model for the Winter Train 1 data set is the PLSRa model. The overflow tonnage for the Winter Train 1 data set is 1086±469 TPH.

Model	Model Performance		PLSRa	
	RMSE	Adj-R^2	RMSE	Adj-R^2
SLR	273 ± 27 TPH	0.05 ± 0.05	114±13 TPH	0.63 ± 0.04
MLR	180 ± 18 TPH	0.59 ± 0.04	15±7 TPH	0.07 ± 0.03
RSR	182 ± 18 TPH	0.59 ± 0.04	17±22 TPH	0.06 ± 0.07
PCR	173 ± 20 TPH	0.59 ± 0.06	16±14 TPH	0.07 ± 0.03
PLSR	176 ± 20 TPH	0.60 ± 0.05	15±8 TPH	0.07 ± 0.03
MLRa	176 ± 11 TPH	0.60 ± 0.05	17±3 TPH	0.07 ± 0.02
PCRa	192 ± 13 TPH	0.53 ± 0.05	33±11 TPH	0.15 ± 0.04
PLSRa	160 ± 10 TPH	0.67 ± 0.04	—	—

Table 4.23: Results for each model for the Winter Train 2 data set. The relative performance between the best performing model and the remaining models is provided in the last two columns. The best performing model for the Winter Train 2 data set is the PLSRa model. The overflow tonnage for the Winter Train 1 data set is 855±279 TPH.

Table 4.23 shows that when only the significant variables were considered, PCR had the best performance with a 37 percent lower RMS error than the SLR model. When all variables were included, the PLSRa model had the best performance of both groups with a 41 percent decrease in RMS error over the SLR model. Note that RSR also performed statistically similar to PLSRa due to large fluctuations in RMS error values resulting in a high standard deviation. In all cases, the regression methods that used the water and geological variables performed significantly better than the feed-only SLR model. When comparing the results for the Winter Train 1 data set and the results for the Winter Train 2 data set, we see again, as in Section 4.5.1, the significant difference in model performance between Train 1 and Train 2 screening systems.

From Tables 4.20 to 4.23, we can see that PLSRa provided the overall best screen model for all data sets. This performance is attributed to its ability to isolate important latent vectors that enable good representation of the underlying system behaviour. This enabled a stronger model to be developed for prediction. In general it was found that using water and geological variables enabled a greater prediction of the screen performance than a simple feed-only model.

4.6 Discussion

In the previous two sections, significant variables were identified for screen operation during summer and winter months. The significant variables were used to develop an explanatory model using multiple linear regression. Several prediction models were then developed using more advanced regression methods. At this time, the significance of the results are discussed.

In Section 4.5.1, significant variables were identified using forward stepwise multiple linear regression for both summer and winter screening conditions. Interestingly, in the summer months geological variables provide a significant portion of the explained variance, however, in the winter period, the water usage variables provided a significant portion of the explained variance. An explanation for this finding is that the cold months freeze the ore transforming it into a rock-like material that is independent of the underlying ore makeup. This supports the observations in the oil sand industry that there exists differences in screening behaviour between summer and winter screening conditions. It was shown that Train 1 and Train 2 behaved similarly during the summer months, however they did not behave similarly in the winter months. As a result, an explanatory screen model required three submodels, a single submodel for summer screening conditions and two separate submodels for winter Train 1 screening conditions and winter Train 2 screening conditions. The coefficients for each of the three screen submodels are listed in Table 4.24.

Summer Screen Model		**Winter Screen Model**		
Variable Name	**Coef.**	**Variable Name**	**Train 1 Coef.**	**Train 2 Coef.**
Feed Tonnage	0.211	Feed Tonnage	0.187	0.077
Pumpbox Valve % Open	19.42	Pumpbox Valve % Open	12.65	14.51
Ore Grade	58.73	Extraction Water Temperature	23.62	33.41
Facies 8	8.72	Flow rate of water to vortex	-1.42	-0.89
Facies 21	-2.62	Recycle Line Water Pressure	-13.35	-7.03
		Pump Gland Water Pressure	-2.42	-0.17

Table 4.24: Variable coefficients for the Summer screen model, the Winter Train 1 screen model, and the Winter Train 2 screen model.

The goal of an explanatory model is to provide insight into the underlying system of interest using the resulting coefficients obtained from multiple linear regression. Meaningful ranking of the variables using the coefficients is difficult due to the complex nature of the collected data. For example, consider the importance of the feed tonnage in the prediction of overflow tonnage. Its ranking should be first because it is directly responsible for the amount of material present on the screen. However, if the screening system is operated in a relatively steady-state manner, the variance of the feed tonnage will be insignificant compared to other variables such as the ore grade which may vary considerably. As a result, the ore grade would be considered a much greater significant variable than the feed tonnage if a ranking system is applied. This results in an inaccurate depiction of the relative importance of each variable. Consequently, a ranking of variables is not considered

until proper designed experiments are conducted.

A second option is to consider the sign of the coefficient of each variable. The sign of the coefficient, or the directional correlation with the response variable, can be used to gain insight into the underlying system. In all three models the feed tonnage is positively correlated with the overflow tonnage. This result is expected since an increase in feed tonnage should cause an increase in overflow tonnage. The second significant variable common to all three models is the pumpbox valve percentage open variable, which was found to be positively correlated with the overflow tonnage. One possible explanation for this result is that as the pumpbox valve closes less water enters the pumpbox and more water is diverted to other areas of the screening system such as the vortex vessel and the spray bars.

The remaining significant variables for the summer period are geological variables. The ore grade variable was found to be positively correlated with the overflow tonnage. Interestingly, this result does not correspond with what is generally understood in the oil industry where increasing the ore grade is found to decrease the overflow tonnage. Facies 8 was found to be negatively correlated with the overflow tonnage and Facies 21 was found to be positively correlated with the overflow tonnage. Descriptions of Facies 8 and Facies 21 are provided in Table 4.25.

Facies	Description	Category
Facies 8	Sand with 10 - 35% clay, clays are in mm to cm thick distinct bedsand dispersed throughout the sand (bioturbated)	Estuarine
Facies 21	Clay/mud with very thin beds (mm) of sand, sand/silt content is 10 - 35%; waste facies	Estuarine

Table 4.25: Facies 8 and 21 descriptions.

From the description of Facies 8 in Table 4.25, an explanation for the negative correlation could be attributed to the presence of large quantities of sand in the material. Unlike Facies 8, Facies 21 contains mud and silt that could increase the level of stickiness in the material, thereby increasing the overflow tonnage.

The remaining significant variables for the winter explanatory regression models are water variables. The water temperature variable was found to be positively correlated with the overflow tonnage. Interestingly, this is opposite to the current understanding at Syncrude where it is observed that the oil sand material residence time on the screens is insufficient for heat transfer from the water to the material. However, since the coefficient of the water temperature variable indicates a positive correlation, is the residence time sufficient for some amount of heat transfer that would affect screen performance? The vortex flow rate variable was found to be negatively correlated with the overflow tonnage. A possible reason for this result is that an increase in flow rate results in a greater amount of material being pushed through the impact screens creating a decrease in overflow tonnage. Similarly the recycle line pressure variable was found to be negatively correlated with the overflow tonnage. A possible explanation is that increasing the recycle line pressure pushes more

water through the line to the vortex. Earlier it was noted that an increase in water flow to the vortex is correlated with a decrease in overflow tonnage, which is now further supported by the results for the recycle line pressure variable. Finally, the pump gland water pressure variable was found to be negatively correlated with the overflow tonnage. A possible explanation is that increasing the pump gland water pressure results in less water to the pump glands and more water the rest of the system such as the vortex. A greater understanding of where the pressure measurements are recorded along the water system is needed for further understanding of the recycle line pressure and the pump gland water pressure variables.

Using the above significant variables, an explanatory model composed of three submodels was developed for summer and winter screening conditions. It was found that a common summer screen submodel increased the RMS error by 2 percent, as measured by the RMS error, over two separate Train 1 and Train 2 screen models. The similar performance of the common and separate summer screen submodels supports the conclusion that Train 1 and Train 2 have similar performance during the summer months. During the winter screening period, it was found that the common screen model increased the RMS error by 12 percent over two separate Train 1 and Train 2 screen models. This result suggests that there exists a difference in behaviour between Train 1 and Train 2 during winter screening conditions. In general, further design of experiment is needed to understand the relationships identified here, however this first step is an important one to identify those variables that have the greatest potential for influence on screen performance.

Following the completion of the explanatory model, predictive models are compared using the identified significant variables. Four linear regression methods, listed in Table 4.26, are considered.

Method	Name
MLR	Multiple Linear Regression
RSR	Response Surface Regression
PCR	Principle Component Regression
PLSR	Partial Least Squares Regression

Table 4.26: List of the four regression methods used to develop the oil sand prediction screen models.

Each method was run a first time on the summer or winter significant variables and then again using all 52 water and geological variables. The only exception is the response surface regression method, which is not run on all 52 variables due to computational restrictions resulting from the large number of predictor variables. Each of the resulting models was evaluated using five-fold cross-validation. Each model was found to perform significantly better than a feed-only model. Among the prediction models, it was found that PLSRa had the best performance with an average 25 percent lower RMS error than the SLR model. In the summer PLSRa decreased the RMS error by 15 percent over the SLR model and 34 percent over the SLR model in the winter. In summary, the results show that using water and geological variables provides greater prediction accuracy of the screen performance than only the feed tonnage.

4.7 Summary

In this chapter, raw data was collected, integrated, and cleaned using assumptions based on typical system operation. The cleaned data was analyzed using forward stepwise multiple linear regression to identify significant variables for both summer and winter operating conditions. It was found that for the summer period, the feed tonnage variable, the ore grade variable, the pumpbox valve percent open variable, the Facies 8 variable, and the Facies 21 variable were significant. For the winter period, the feed tonnage variable, the pumpbox valve percent open variable, the recycle line pressure variable, the water temperature variable, the gland pump pressure variable, and the vortex flow rate variable were significant. Using the identified significant variables, an explanatory model composed of one submodel for the summer period and two submodels for the winter period was developed. Four regression methods were used to create predictive models using water and geological variables. It was found that a PLSR model that uses all 52 water and geological variables performed the best with an average 25 percent decrease in RMS error over a simple feed-only model. This shows that substantial improvement in system description and system performance prediction can be achieved through the inclusion of water and geological variables.

Chapter 5

Conclusion and Future Work

5.1 Oil Sand Screen Modelling

Through the analysis of geological and system water data, a new model is developed for oil sand screens in operation at Syncrude Canada Ltd. The model is composed of one submodel for summer screening conditions and two submodels for winter screening conditions. The submodels for winter screening conditions take into account unexplained differences found between Train 1 and Train 2 screening performance. Several statistical measures were used to ensure that each submodel was statistically significant.

An interesting result from the formation of the submodels was that summer screening performance is affected by primarily geological variables, and winter screening performance is affected primarily by system water variables. Summer screen variables include the feed tonnage, the ore grade, the pumpbox valve percent open, the Facies 8, and the Facies 21 variables. Winter screen variables include the feed tonnage, the pumpbox valve percent open, the recycle line pressure, the water temperature, the gland pump pressure, and the vortex flow rate variables. However, the gland pump pressure variable was found to be significant for only Train 2 data. The identified significant variables validate claims made by operators concerning their significance.

Seven separate prediction models developed using four linear regression methods were evaluated using five-fold cross-validation. Each model was shown to perform significantly better than a model that considered only the feed tonnage as input variable. Of the seven models, a partial least squares regression model using all 52 significant variables performed the best with an average 25 percent reduction in RMS error over a simple feed-only model.

A number of research contributions are made in this work. We have shown that geological and system water variables are important in modelling oil sand screens. A preliminary screen model was developed and shown to provide significant additional explained variance over a simple feed-only screen model. Furthermore, important geological and system water variables were identified for both summer and winter screening conditions that provide direction for future studies.

The overall relative increase in performance of the new screen model over that of the previous

feed-only model is significant, particularly for winter screening conditions. However, the performance was limited in part due to the lack of time series feed particle size distribution information. Consequently, the overall absolute explained variance remains low, particularly for summer screening conditions, which motivates the need for future work in the area of oil sand screen modelling.

5.2 Future Work

The historical nature of the data limits the conclusions made in this work to those situations observed by the data. Consequently, it is important to understand that the findings of the regression methods are only correlational and not causal. Further analysis carried out in a controlled process is an essential step towards validating the relationships proposed in this work. We now present several directions for future work in oil sand screen modelling.

Previous vibrating screen modelling research focused largely on the feed particle size distribution which was not available at the time of this work. This limited the applicability of previous research work and the existing knowledge of screen behaviour that may have been transferable to oil sand static screens. Future work should involve detailed collection of the feed particle size distribution.

The accuracy of the sensor data collected throughout the plant is largely unknown and affects the quality of the regression results and subsequent conclusions. To increase the accuracy of the data analysis, the measurement error should be evaluated and ultimately reduced if it is found that the accuracy of the measurements is not within reasonable limits.

In this work, it was determined that unexplained differences between Train 1 and Train 2 screen performance existed. Further investigation into the reasons for this would require the collection of additional screen variables not considered in this work. A greater understanding of the screening performance difference would lead to better operating practices for both screening systems.

The work presented here addresses summer and winter screening behaviour. Future work should address the development of spring and fall screening models so that a model for the entire year is developed.

The future work outlined in this section would enhance the results obtained thus far and provide a stronger understanding of the oil sand static screening system in use at Syncrude's North Mine.

Bibliography

[1] A. Apling. Measuring screen performances. *Mine and Quarry*, 14(4):31–35, April 1985.

[2] A. Apling, R. Wang, and T. Bereton. An investigation into factors affecting screen performance. *5th International Mineral Processing Symposium*, pages 39–48, September 1994.

[3] Environment Canada National Climate Archive. website, 2006. `http://www.climate.weatheroffice.ec.gc.ca/`.

[4] R.J. Batterham, K.R. Weller, T.E. Norgate, and C.J. Birkett. Screen performance and modelling with special reference to iron ore crushing plants. *European Symposium of Particle Technology*, pages 842–857, 1980.

[5] J.M. Beeckmans and A. Jutan. Monte carlo simulation of a probability screen. *Canadian Journal of Chemical Engineering*, 67:329–336, Apr. 1989.

[6] S. Bhattacharya and D.D. Misra. Computer simulation of a screening circuit. *Chemical Engineering World*, 23:56–61, May. 1988.

[7] M. Bothwell and A. Mular. Coarse screening. In A. Mular, D. Halbe, and D. Barratt, editors, *Mineral Processing Plant Design, Practice, and Control*, pages 894–916. Society of Mining, Metallurgy, and Exploration, Inc, 2002.

[8] G. Box, W. Hunter, and J. Hunter. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. Robert Maxwell, M.C., John Wiley & Sons, Inc, first edition, 1978.

[9] T. Brereton. Performance measurement of a screening process using statistical techinques. *Quarry Managers Journal*, 54:344–346, 1970.

[10] T. Brereton and K.R. Dymott. Some factors which influence screen performance. *10th International Mineral Processing Congress (1973)*, pages 181–194, 1974.

[11] W. Brown and K. Wohletz. Derivation of the weibull distribution based on physical principles and its connection to the rosin-rammler and log-normal distributions. *Journal of Applied Physics*, 78:2758–2763, 1995.

[12] J.M. Beeckmans Chen Rongguang and Chen Qingru. A convenient correlation for modelling the performance of probability screens. *International Journal of Mineral Processing*, 36:31–40, 1992.

[13] K. Colman. *Selection Guidelines for Vibrating screens*, chapter 2 section 3E, pages 13 – 19. SME Mineral Processing Handbook. SME, Littleton, CO, 1982.

[14] Personal communication with Brenda Wright. Syncrude mine planning. 2006.

[15] H. Crissman. Vibrating screen selection. *Pit and Quarry*, 78:39–40,42,44, June 1986.

[16] R. Feller and A. Foux. Oscillating screen motion effect on particle passage through perforations. *Transactions of the ASAE*, 18:926–931, 1975.

[17] G. Ferrara and U. Preti. A contribution to screen kinetics. *Proc 11th Int Miner Proc Congress*, pages 183–217, 1975.

[18] G. Ferrara and G.D. Schena. Modelling and simulation of integrated plant operations of mineral processing. *Control '84*, pages 153–165, 1984.

[19] A.M. Gaudin. *Principles of Mineral Dressing*. McGraw-Hill, New York, 1939.

[20] S.E. Gluck. Vibrating screens: Surface selection and capacity. *Chemical Engineering*, 72:179–184, Mar. 1965.

[21] C.C. Hatch and A.L. Mular. Simulation of the brenda mines ltd secondary crushing plant. *SME-AIME Annual General Meeting*, 1979.

[22] F.W. Hess. *Mathematical Modeling of Screens and Related Units for Plant Simulation*. Phd thesis, University of Queensland (JKMRC), 1983.

[23] C.L. Karr and B. Weck. Computer modelling of mineral processing equipment using fuzzy mathematics. *Minerals Engineering*, 9(2):183–194, 1995.

[24] V.K. Karra. Development of a model for predicting the screening performance of a vibrating screen. *Cim Bulletin*, pages 167–171, 1979.

[25] E. Kelly and D. Spottiswood. *Introduction to Mineral Processing*. Wiley, New York, USA, 1982.

[26] R.P. King. *Modeling and Simulation of Mineral Processing Systems*. Butterworth-Heinemann, 2001.

[27] Syncrude Canada Ltd. website, 2006. `http://www.syncrude.ca/`.

[28] A.J. Lynch. *Mineral crushing and grinding circuits. Their simulation, optimsation, design, and control*. Elsevier Scientific Pub. Co., Amsterdam, New York, 1977.

[29] H. Martens and T. Naes. *Multivariate Calibration*. Biddles Ltd., Guildford, 1991.

[30] R. Mason, R. Gunst, and J. Hess. *Statistical Design and Analysis of Experiments: With Applications to Engineering and Science*. John Wiley & Sons, first edition, 1989.

[31] C.W. Matthews. Trends in size reduction of solids - screening. *Chemical Engineering*, 79:76–83, Jul. 1972.

[32] S. Miwa. Proposal of a new index for expressing the performance of screens. *Kagaku Kogaku*, 24:150–153, 1960.

[33] J. Mosby. *Investigation of the segregation of particulate solids with emphasis on the use of segregation testers*. PhD thesis, Telemark College, Norway, 1996.

[34] A. Mular. *Principles of Mineral Processing*, chapter 4. Society for Mining, metallurgy, and Exploration, Inc., Littleton, Colorado, 2003.

[35] R. Myers. *Cassical and Modern Regression with Applications*. PWS Publishers, 1986.

[36] T.J. Napier-Munn, S. Morrell, R.D. Morrison, and T. Kojovic. *Mineral Comminution Circuits - Their Operation and Optimisation*. Julius Kruttschnitt Mineral Research Centre, Queensland, Australia, 1999.

[37] J. Neter, M.H. Kutner, C.J. Nachtsheim, and W. Wasserman. *Applied Linear Regression Models*. Chicago, Ill. : Irwin, c1996., 1996.

[38] J. Neter, W. Wasserman, and M.H. Kutner. *Applied linear statistical models*. Irwin, Boston, 1990.

[39] J. Nichols. *Selection and Sizing of Screens*, chapter 27, pages 509–522. Design and installation of comminution circuits. SME, Littleton, CO, 1982.

[40] C. Oshins, K. Fenstermacher, and S. Bush. How much does it cost to run a country yard trimmings recycling system. *BioCycle*, 46(3):45–51, 2005.

[41] Inc. OSIsoft. website, 2007. `http://www.osisoft.com/`.

[42] A.C. Partridge and J. Roberts. Principles of screening. *Mine and Quarry*, pages 33–38, December, 1977.

[43] A.N. Pritchard. Vibrating screens in the mining industry. *Mine and Quarry*, 9:46, Oct. 1980.

[44] H.G. Reading. *Sedimentary Environments and Facies*. Blackwell Scientific Publications, 1996.

[45] P. Rosin and E. Rammler. Laws governing the fineness of powdered coal. *J. Inst. Fuel*, 7:89–105, 1933.

[46] D. Sandoval. Building a niche. *Recycling Today*, 43(1):22–26, 2005.

[47] R. Schuhmann. Principles of comminution, 1-size distribution and surface calculation. *AIME Techincal Publications*, 1940.

[48] Surpac Software. website, 2006. `http://www.surpac.com/`.

[49] M. Soldinger. Interrelation of stratification and passage in the screening process. *Minerals Engineering*, 12(5):497–516, 1999.

[50] M. Soldinger Stafhammar. *Screening of crushed rock material*. Ph.d. dissertation, Chalmers Tekniska Hogskola, Sweden, 2002.

[51] Standish, Bharadwaj, and Hariri-Akbari. A study of the effect of operating variables on the efficiency of a vibrating screen. *Powder Technology*, 48:161–172, 1986.

[52] Standish and Meta. Some kinetic aspects of continuous screening. *Powder Technology*, pages 165–171, 1985.

[53] G. Subasinghe, W. Schaap, and E. Kelly. Modelling the screening process - a probabilistic approach. *Powder Technology*, 59:37–44, 1989.

[54] G. Subasinghe, W. Schaap, and E. Kelly. Modelling the screening process - an empirical approach. *Minerals Engineering*, 2(2):235–244, 1989.

[55] G.K.N.S. Subasinghe, W. Schaap, and E.G. Kelly. Modelling screening as a conjugate rate process. *International Journal of Mineral Processing*, 28:289–300, 1990.

[56] F.M. Sultanbawa, W.G. Owens, and S.S. Pandiella. A new approach to the prediction of particle separation by sieving in flour milling. *Trans iChemE*, 79(Part C):211–218, December 2001.

[57] J.S.J. Van Deventer T.J. Van der Walt and E. Barnard. Neural nets for simulation of mineral processing operations: Part i. theoretical principles. *Minerals Engineering*, 6(11):1127–1134, 1993.

[58] J.S.J. Van Deventer T.J. Van der Walt and E. Barnard. Neural nets for simulation of mineral processing operations: Part ii. applications. *Minerals Engineering*, 6(11):1135–1153, 1993.

[59] W.J. Whiten. The simulation of crushing plants with models developed using multiple spline regression. *10th International Symposium on the Application of Computer Methods in the Mining Industry*, 72:257–264, 1972.

[60] W.J. Whiten. Models and control techniques for crushing plants. *Control 84, Mineral Metalurgical Processing*, pages 217–224, 1984.

[61] W.J. Whiten and M.E. White. Modelling and simulation of high tonnage crushing plants. *12th International Mineral Processing Congress*, 2:148–158, 1977.

[62] B. Wills. *Industrial Screening*, volume 41 of *Mineral Processing Technology*, chapter 8, pages 309–334. Robert Maxwell, M.C., Pergamon Press, fourth edition, 1988.

[63] Y. Zhao. Complicated motion of particles on vibration screen surface. *Transactions of Nonferrous Metals Society of China*, 9(4):861–864, 1999.